Barigozzi, Matteo; Cuzzola, Angelo; Grazzi, Marco; Moschella, Daniele

Working Paper

Factoring in the micro: A transaction-level dynamic factor approach to the decomposition of export volatility

LEM Working Paper Series, No. 2021/22

Provided in Cooperation with:
Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies

Suggested Citation: Barigozzi, Matteo; Cuzzola, Angelo; Grazzi, Marco; Moschella, Daniele (2021) : Factoring in the micro: A transaction-level dynamic factor approach to the decomposition of export volatility, LEM Working Paper Series, No. 2021/22, Scuola Superiore Sant'Anna, Laboratory of Economics and Management (LEM), Pisa

This Version is available at:
https://hdl.handle.net/10419/243518

INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

Factoring in the micro: a transaction-level dynamic factor approach to the decomposition of export volatility

Matteo Barigozzi [a]
Angelo Cuzzola [b]
Marco Grazzi [c]
Daniele Moschella [b]

[a] Department of Economics, Università di Bologna, Italy.
[b] Institute of Economics and Department EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy.
[c] Department of Economic Policy, Università Cattolica del Sacro Cuore, Milan, Italy.

2021/22 **June 2021**
ISSN(ONLINE) 2284-0400

Factoring in the micro: a transaction-level dynamic factor approach to the decomposition of export volatility*

Matteo Barigozzi[1], Angelo Cuzzola[2], Marco Grazzi[3], and Daniele Moschella[2]

[1]Department of Economics, Università di Bologna
[3]Department of Economic Policy, Università Cattolica del Sacro Cuore, Milano (Italy)
[2]Institute of Economics & Department EMbeDS, Scuola Superiore Sant'Anna, Pisa, (Italy)

June 8, 2021

Abstract

This paper analyzes the sources of export volatility estimating a dynamic factor model on transaction-level data. Using an exhaustive dataset covering all French export transactions over the period 1993-2017, we reconstruct the latent factor space associated to global and destination-specific macroeconomic cycles by means of a modified expectation maximization algorithm to accommodate both the sparsity and the high dimensionality of the micro time series. Thus while paving the way for a novel application of dynamic factor models to microeconomic analysis, we provide a decomposition of the volatility of aggregate export and firms growth rates, highlighting structural spatial patterns and drawing attention to the role of geographical diversification for the mitigation of risks related to firms' export activities.

Keywords: Factor models, trade volatility, diversification
JEL classification: C38, L25, F14

*This paper is the result of a long process that over time, at different stages, benefited from helpful comments at several conferences. We are also indebted to Julien Martin and Marco Cococcioni, for insightful comments. This work has been partly supported by the European Commission under the H2020, GROWINPRO, Grant Agreement 822781 and by Italian Ministry of Education and Research under the PRIN-2017 Programme, project code 201799ZJSN. This work is also supported by a public grant overseen by the French National Research Agency (ANR) as part of the 'Investissements d'avenir' program (reference: ANR-10-EQPX-17, Centre d'accès sécurisé aux données, CASD). The usual disclaimer applies.

1 Introduction

A simple inspection of the components of the GDP reveals that export stands out as the most volatile part and that trade openness contributes to increase aggregate volatility (see di Giovanni and Levchenko, 2009 and Figure 1a). This stylized fact is particularly relevant to the characterization of the risks associated to international trade along the growth paths of countries, sectors and firms, and justifies the recent increasing interest on the relationship between trade and income volatility (Caselli et al., 2020).

Granted this, a fundamental question revolves around the sources of aggregate trade fluctuations. In this perspective, a first wave of research has investigated to what extent general aggregate shocks come from more fundamental, sectoral shocks, and to which degree sectoral diversifcation may reduce aggreagate volatility (Long Jr and Plosser, 1983; Koren and Tenreyro, 2007). Starting from the seminal work by Gabaix (2011), a rich stream of literature has then shown that in a granular economy idiosyncratic shocks to individual firms may lead to large aggregate movements (Acemoglu et al., 2012; Carvalho and Gabaix, 2013; Carvalho and Grassi, 2019). Taking such a viewpoint, if larger firms are expected to contribute to generate large shocks in the economy, this should be particularly true in international trade, where the firm size distribution is even more skewed (among the others, Bernard et al., 2009, 2016). In such a context, the exposure to international shocks might further increase the risk of granular shock (di Giovanni and Levchenko, 2009; di Giovanni et al., 2014) as well as induce international business cycle comovement (Di Giovanni et al., 2018).

Following this line of investigation, our work proposes a novel approach to the identification and quantification of the sources of export volatility at different level of aggregation, estimating a dynamic factor model on the growth rates of export flows[1] for the universe of French exporting firms.

We propose a detailed characterization of the macroeconomic components, encoding in the model the dynamics of the macroeconomic factors to estimate common shocks. Moving from transaction-level, quarterly data we are able to estimate dynamic effects on a sufficiently long time series and at the same time we are able to mitigate the potential bias that partial-year effects may have in capturing exporter dynamics (Bernard et al., 2017). We perform these exercises using data from 1993 to 2017: this span includes relevant macroeconomic events (i.e. the trade collapse) which might be relevant in the estimated contribution to aggregate volatility of common movements.

We contribute to several streams of literature. *First,* we provide a closer look at the granularity hypothesis with a model designed to capture the relevance of global and destination-specific export shocks, not only *per se* but also as drivers of heterogeneous responses at the firm level. *Second,* we decompose the volatility at the firm level providing new insights on the volatility-diversification nexus. Several contributions find a dampening effect of diversification on volatility (Bottazzi and Secchi, 2006; Kelly et al., forthcoming); in particular, firms selling multiple products to multiple countries, which make most of aggregate trade flows (Eaton et al., 2004; Bernard et al., 2012), could reduce their volatility by diversifying their portfolio (di Giovanni et al., 2014; Kramarz et al., 2020). Our decomposition allows us to to identify which

[1] As will be clear in the following, export flows are intended as firm-destination pairs.

share of risks can be diversified away.

Once estimated the different components of flows' growth rates, we recover the volatility decomposition at the aggregate and firm-level of analysis. On the former, our results confirm only in part the general wisdom recognising the dominance of idiosyncratic shocks in originating aggregate volatility: its measured impact results reduced with respect to the estimates of Gabaix (2011); Carvalho and Gabaix (2013); di Giovanni et al. (2014), favouring a slight increase of the influence of common macroeconomic components. This is possibly due to the peculiar structure of the model that postulates the existence of latent factors and of heterogenous flow-level loadings. In fact, global and destination-specific shocks are defined as the interplay between a latent common factor and flow-specific coefficients, the latter capturing the variation of single flows due to the idiosyncratic response to common shocks.

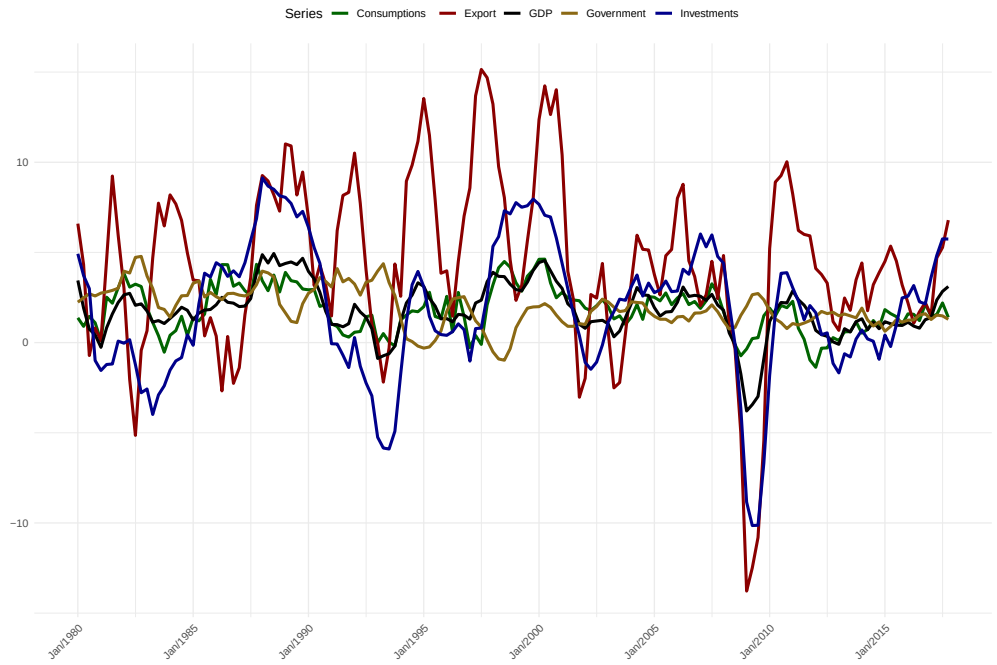
At the micro-level, we measure the effects of the diverse sources on firms' volatility distribution, showing how global and destination specific components originate a significant part of the risks inherent to export growth, even though the impact of the idiosyncratic non reducible components is relatively higher. The same decomposition is then used to understand how and to what extent geographical diversification strategies help dampening both export volatility overall and the single components. We find that while firms diversifying across different destinations succeed in mitigating the risks associated to the macroeconomic cycle, a reverse U-shaped relation between diversification and idiosyncratic volatility suggests that the same strategies do not dampen idiosyncratic risks until a certain level of diversification is reached.

The remainder of the paper is structured as follows. Section 2 introduces the model and gives a brief and concise description of the methodology. Section 3 offers a bird's eye view on the dataset characteristics, with a particular focus on sparsity and some firm-level statistics. Section 4 consists of a technical description of the quasi-maximum likelihood estimation method for dynamic factor models with arbitrary pattern of missing data. Section 5 presents the results: the reconstruction of the latent factor space and the volatility decomposition at the aggregate and firm-level. Section 5 concludes.

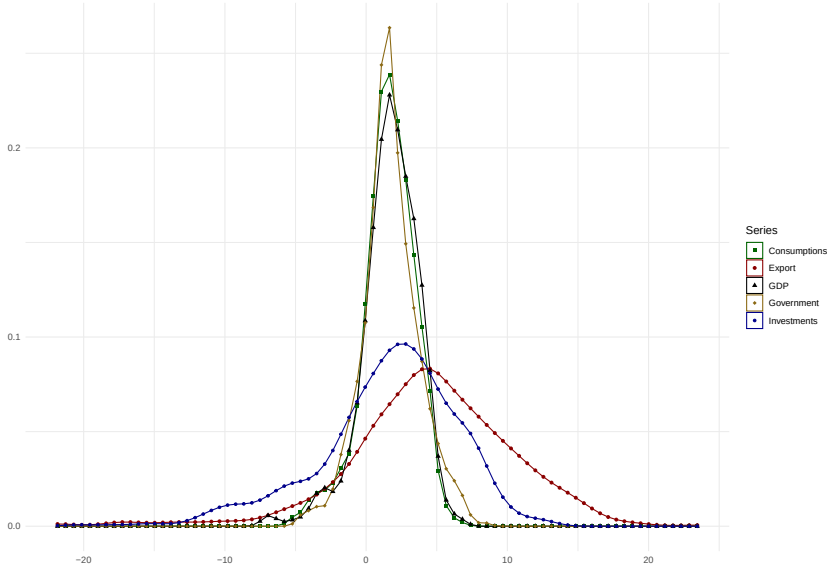
2 Model and methodology

In this work we provide an econometric framework to identify the contribution of different sources of shocks affecting international trade flows. As described in the following, the methodology guarantees a high degree of flexibility as the structure that we impose *a priori* is kept to a minimum and can ultimately be referred to the nature of the disaggregated transaction data.

In brief, the approach allows for the identification of global and destination-specific shocks hitting international export flows of firms and their influence on aggregate and firm-level growth patterns. In so doing we build upon the framework laid down by di Giovanni et al. (2014) and Kramarz et al. (2020) where the authors decompose micro-level growth rates in terms of orthogonal shocks, of both macroeconomic and microeconomic nature. We share the general baseline of the orthogonal decomposition, yet proposing a more detailed characterization of the macroeconomic components, encoding in the model the dynamics of the macroeconomic factors and a flow-specific



(a) Time-series.



(b) Kernel densities.

Figure 1: Our elaboration on the FRED data offering a comparison of the growth rates of the main GDP components. Figure (a): time series of the quarterly deseasonalized growth rates for France. Figure (b): Gaussian kernel estimates of the pooled distribution of quarterly GRs for France, Italy, Germany and Spain from 1993 to 2018.

response to the identified common shocks. The flow's growth rate is then recovered as the sum of three parts. The first two terms identify the flow-specific response to dy-

namic factors influencing at any time step i) all the transactions in the dataset (global factor), ii) only the flows directed to a specific destination (destination-specific factors). The third component is instead an idiosyncratic term, absorbing any non-reducible flow-specific effects on the growth rate. Technically, this consists in estimating a dynamic factors model on flows' growth rates with a block structure induced by spatial diversification patterns. In fact, the chosen methodology not only allows to recover the latent factor space of shocks influencing international trade, but provides an estimation of the idiosyncratic responses as flow-specific factor loadings. Once estimated, the parameters of the model serve as the baseline for flow's growth rate decomposition that is then scaled up to firm and total aggregate level through weighted aggregation.

Figure 2 provides a schematic description of the information flow and the implied dynamics for the outlined framework. The proposed methodology shows remarkable differences with respect to the most common orthogonal decomposition models (ODMs): dynamic factor models convey the information of the sequential cross-sections chaining export flows co-movements into a latent factor that follows a given stochastic process, while providing a characterization of the idiosyncratic responses.

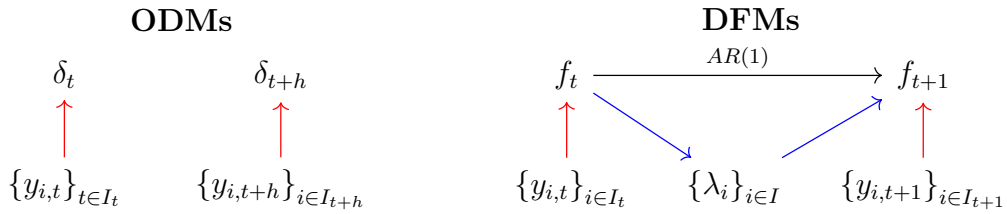


Figure 2: Comparing orthogonal decomposition models (ODMs) and dynamic factor models (DFMs): the former's class estimates the components as series of independent cross-section statistics, whereas the latter summarizes the information coming from the cross-section into latent factors whose dynamics is encoded in a general stochastic framework (e.g. $AR(p)$). In the vertical dimension information flows are bidirectional: in a two-step estimation procedure, the estimated factors determine the idiosyncratic factor loadings that are used in the following step to update the factor estimates until convergence. The interplay between the horizontal and vertical dimensions allows a full fledged dynamic decomposition and a most efficient handling of the available information.

Model's equations Formally, on destination exporter pairs the model is described by the set of equations:

$$y_{de,t} = \lambda_{de}\, f_t + \rho_{de}\, g_{d,t} + \xi_{de,t} \tag{1}$$
$$f_t = a_f\, f_{t-1} + u_{f,t} \tag{2}$$
$$g_{d,t} = a_d\, g_{d,t-1} + u_{d,t} \tag{3}$$

and postulates the influence on the flow-level growth rate ($y_{de,t}$) of a latent factor common to all the flows, f_t, and one $g_{d,t}$ specific to the destination d. The loadings λ_{de} and ρ_{de} model the specific response to shocks affecting the related factors which come endowed with some additional conditions on their dynamics, encoded in independent autoregressive processes of order one[2]. Relying on the vectorized short hand notation

[2] Estimation of the model with AR processes of higher orders is possible and technically feasible. Nevertheless, we have chosen a very simple picture to avoid proliferation of parameters and potential overfitting.

(e.g. $F = (f_1, \ldots, f_T)'$, the matrix formulation highlights the block structure imposed by the model

$$\begin{pmatrix} Y_1 \\ \vdots \\ \vdots \\ Y_D \end{pmatrix} = \begin{pmatrix} \Lambda_1 & R_1 & 0 & \cdots & 0 \\ \vdots & 0 & R_2 & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Lambda_D & 0 & 0 & \cdots & R_D \end{pmatrix} \cdot \begin{pmatrix} F \\ G_1 \\ \vdots \\ G_D \end{pmatrix}$$

Within this framework, the factor F impacts — and is estimated out of — all the considered flows through $\Lambda_d = (\lambda_{de_1}, \ldots, \lambda_{de_{n_D}})$, whereas the remaining factors track destination-specific comovements that are residual to the common component but relevant to the destination country only. In other words, the matrix equation for a generic block, $Y_d = \Lambda_d F + R_d G_d$, captures the dynamics of the flows belonging to a particular block as first order interaction with a factor commmon to all the export flows and an orthogonal second order block-specific factor. This implicitly defines a hierarchy of the factor and limit the application of the method to model designs admitting a pecking order with meaningful economic assumptions.

Empirical estimation Dynamic factor models (DFM) have an established tradition in macroeconometrics. Their early applications were based on the assumption that all cross-correlations in the data could depend on few common factors while the idiosyncratic noise remain cross-sectionally uncorrelated. Unfortunately, this becomes an unrealistic hypothesis in the case of interest to our research question, where the dataset's cross-sections is very large, favouring the emergence of cross-sectional correlations. For this reason, empirical estimation is based on the so called approximate DFMs, recently developed to provide consistent estimates in presence of limited correlation among idiosyncratic components (see e.g. Forni et al., 2000; Stock and Watson, 2002, for classical principal component approach). Within this stream of literature, more recent works by Doz et al. (2012); Bańbura and Modugno (2014) introduce the so called quasi-maximum likelihood (QML) method which allow to recover consistent estimates of the components as parameters of the maximum likelihood of a miss-specified factor model, achieving two crucial improvements in our context: i) the level of complexity is anchored to the number of factors and not to number of series, ii) QML approach can be adapted to deal with missing data and *ad hoc* defined block structures imposed by the model. Building upon these contributions, we further extend such framework. *First*, aiming at disentangling global and destination-specific effects we model a block structure imposing *ad hoc* zero restrictions on the factor loadings. This consists in estimating destination-specific shocks as factors that are common to blocks of series recovering the complete structure of the parameters afterwards (see the technical exposition below for further details). *Second*, we optimize the base algorithm and propose an inizialization based on (Breitung and Eickmeier, 2015) to deal with a dataset with very large cross-sections, approaching an order of magnitude of $\sim 10^6$ time series. In this respect, notice that, for the sake of comparison, standard applications of dynamic factor models are estimated on few hundreds of time series. *Third*, we work with heterogeneous patterns of missing data, yet obtaining successful results in terms of factor identification even with a sparsity ratio exceeding the 80%. The proposed approach consist of an iterated estimation procedure called Expectation

Maximization (EM) algorithm which converges towards the maximum likelihood estimates of the model in a sequence of steps. Denoting in short, the factors as Z, the other parameters[3] with θ and the matrix of data with Y, we write the log-likelihood, $l(Y, Z, \theta)$, associated to the system ((1), (2) and (3)) as a function according to the following steps[4]:

1. Given the factors, the parameters are derived by analytical log-likelihood maximization. The expected log-likelihood is computed conditional on the available information and parameters estimate of the previous iteration

$$L(\theta) = \mathrm{E}_\theta \left[l\left(Y, Z^{(k-1)}, \theta\right) | \Omega_T \right] \quad .$$

   Updated parameters, $\theta^{(k)}$ are analytically derived maximizing $L(\theta)$.

2. Given the parameters, the factors are derived through Kalman Filter/Smoother iteration. Once $\theta^{(k)}$ has been updated, $F^{(k)}$ matrix is computed running the Kalman Smoother on the state space model defined by the equations above, taking the conditional expectation of the true factors given the estimates of θ.

This cycle defines a sequence of increasing log-likelihood values

$$l\left(Y, Z^{(0)}, \theta^{(0)}\right) \rightarrow l\left(Y, Z^{(0)}, \theta^{(1)}\right) \rightarrow l\left(Y, Z^{(1)}, \theta^{(1)}\right)$$

and stops when an appropriate convergence condition is fulfilled. Notice that the analytical expression for the log-likelihood is derived in explicit form assuming gaussianity of the idiosyncratic components and of the factor's innovation. In fact, theoretical results on the consistency and the efficiency of the method are derived under gaussianity. Nonetheless, several empirical and numerical analysis find that the estimates are robust when the distribution of idiosyncratic terms is non-gaussian and displays fatter and asymmetric tails (see e.g. Reis and Watson, 2010; Barigozzi and Luciani, 2019), suggesting that this methodology, once adapted, can be applied to growth rates of export transactions whose distributional properties will be explored in the following section.

In order to adapt the EM algorithm to our dataset, we optimize its computational implementation for the repeated estimation on random samples of the original dataset. In practice, we first build a predefined number of partial dataset selecting at random a fixed number of firms (avoiding possible biases in the destinations representation), then we apply the estimation procedure to each reduced dataset and recover the factors and the aggregate volatility estimates by simple averages, constructing the relative confidence intervals. Analogously, we control that the firm level volatility distributions are constant across samples and exploit this result to decompose the volatility at the microeconomic level. This approach reduces the computational overload on the key steps of the EM algorithm and allows to check the robustness of the model to diverse sample biases.

[3] Namely, the loadings matrix and relevant variance covariance matrices. See the technical details in the related section below.

[4] Index with parenthesis run over the sequential estimation steps.

Volatility decomposition Once estimated, the model master equation provides a flow level decomposition of the logarithmic growth rates

$$\hat{y}_{de,t} = \hat{\lambda}_{de}\,\hat{f}_t + \hat{\rho}_{de}\,\hat{g}_{d,t} + \hat{\xi}_{de,t}$$

Moving from that, the decomposed flows' grow rates can be mapped to different aggregates. Thus we analyze both from a macroeconomic aggregates, estimating the volatility components of the total export and of the export to specific destinations[5], and microeconomic ones, focusing on firm level volatilities. When comparing the growth rate of the aggregate with the rates of its parts it is not always possible to get an exact relation and often one has to search for reliable proxies, making the micro-to-macro mapping non-canonical and highly dependent on the properties of the data and the nature of the research question[6]. In the following, we offer a comparison between standard aggregation obtained through weighted sum of the flows (see e.g. Gabaix, 2011; di Giovanni and Levchenko, 2009; Kramarz et al., 2020, relying on constant or dyamic weigths) and some recently developed method accounting explicitly for distributional effects (Bottazzi et al., 2019) (Appendix B presents the methods together with a discussion of the possible limits). From a general perspective, all these methods are based on the construction of a function that is linear in the flows' components (not necessarily in the space of the flows' time series). Formally, denoting by Agg the generic aggregation strategy, one is induced to consider an equation of the form

$$\underbrace{\mathrm{Agg}\left(\left\{\hat{\lambda}_{de}\,\hat{f}_t\right\}\right)}_{Glob_t} + \underbrace{\mathrm{Agg}\left(\{\hat{\rho}_{de}\,\hat{g}_{d,t}\}\right)}_{\mathrm{Dest}_t} + \underbrace{\mathrm{Agg}\left(\left\{\hat{\xi}_{de,t}\right\}\right)}_{\mathrm{Idio}_t}$$

whereby the three time series can be interpreted as, respectively, the global/common, destination-specific and idiosyncratic/granular components. Then, the variance (or standard deviation) of the series gives a proxy of the volatility that is associated to a single specific component. As an example, one can recover the aggregate decomposition after defining the weights[7]:

$$\omega_{i,t}^{agg} = \frac{y_{i,t-1}}{\sum_j y_{j,t-1}}\,.$$

and write the following volatility decomposition:

$$\begin{aligned}\mathrm{Var}_T\left(\sum_i \omega_{i,t}^{agg} y_{i,t}\right) = \;& \mathrm{Var}_T\left(\textstyle\sum_i \omega_{i,t}^{agg}\cdot\lambda_i\cdot f_t\right) + \mathrm{Var}_T\left(\textstyle\sum_i \omega_{i,t}^{agg}\cdot\sum_c \rho_{c,i}\cdot d_{c,t}\right) \;+ \\ & \mathrm{Var}_T\left(\textstyle\sum_i \omega_{i,t}^{agg}\cdot\xi_{i,t}\right) + \mathrm{Cov} \qquad (4)\end{aligned}$$

Moreover, with the aim of studying the volatility associated to a specific destination one can both restrict to the country-specific subset of the flows (I_c) and look at the

[5] With a minimum impact on the analysis a reduction of the dataset is necessary for the practical implementation of the estimation method. In the following, we will anyway refer to 'total export' as the entirety of the dataset after restrictions.

[6] Not many papers explore thoroughly the possible implications of this trivial fact. A remarkable exception is the work of Amiti and Weinstein (2018) where, within a firm-bank orthogonal decomposition model, the authors show how to adapt the aggregation strategies to estimation techniques adopted to identify the micro-level shocks.

[7] Here, until the end of the paragraph, we simplify the notiation using i or j as multinidices running over exporter-destination pairs (e, d)

variance of the global, of the destination-specific components or at the sum of the two. After rescaling the weights

$$\omega_{i,t}^{c} = \frac{y_{i,t-1}}{\sum_{j \in I_c} y_{j,t-1}} \, .$$

$$\mathrm{Var}_T \left(\sum_{i \in I_c} \omega_{i,t}^{c} \cdot \rho_{c,i} \cdot d_{c,t} \right) \tag{5}$$

Aggregation through dynamic weighting applies also to the firm-level aggregation, once recovered the firm's portfolio (I_s) and defined the firm-level weights

$$\omega_{i,t}^{(s)} = \frac{y_{i,t-1}}{\sum_{j \in I_s} y_{j,t-1}} \, .$$

to get eventually the volatility proxy:

$$\begin{aligned} \mathrm{Var}_T \left(\sum_{i \in I_s} \omega_{i,t}^{(s)} y_{i,t} \right) = \; & \mathrm{Var}_T \left(\textstyle\sum_{i \in I_s} \omega_{i,t}^{(s)} \cdot \lambda_i \cdot f_t \right) + \mathrm{Var}_T \left(\textstyle\sum_{i \in I_s} \widetilde{\omega}_{i,t}^{(s)} \cdot \sum_c \lambda_{c,i}^{(d)} \cdot d_{c,t} \right) \; + \\ & \mathrm{Var}_T \left(\textstyle\sum_{i \in I_s} \omega_{i,t}^{(s)} \cdot \xi_{i,t} \right) + \mathrm{Cov} \end{aligned} \tag{6}$$

whose components are again interpreted according to the classification given before. For this last aggregation step, it is not possible to rely on the aggregation methods as proposed in Appendix B: indeed the relevant empirical distributions have few points to be estimated out of because i) many firms do not report enough country level flows (many exporters trade with partners in limited number of countries), ii) even restricting the analysis to very diversified exporters only, the number of destination considered in the analysis defines an upper bound of less than one hundred of observations.

3 Data and stylized facts

To estimate the outlined model, we rely on transaction-level exports recorded by the French customs office (*Direction Générale des Douanes et des Droits Indirects*, DGDDI).[8] The dataset contains detailed information on export flows on monthly basis for each year from 1993 to 2017 for all French exporters. Each exporter is identified by a unique official identification number (SIREN code) and transactions report export value, quantity, country of destination, and an 8-digit product code following the European Union's Combined Nomenclature (CN8). We start by applying standard cleaning methodologies described in Bergounhon et al. (2018) and often employed on this set of data. They consist in the harmonization of product codes, constructing a coherent chain of HS system's labels, and homogenization of registered transactions. As to the latter, since the registration of the transactions below the threshold of 1000 euros (or 1000 Kgs) was not compulsory before 2010, we opted for the deletion of all the transactions not reaching the threshold before and after 2010. The number of involved elements is consistent, around 1.5 millions of firm-product-destination-month tetrads

[8] The data are directly provided to researchers by the DGDDI upon the approval of a research proposal by the *Comité du Secret Statistique*.

per year, yet accounting for a very small fraction the total export value (around 0.5%). This basic cleaning leaves an average value of export per year of euros 340.99 billions and, after aggregating along the product dimension, 3.2 millions of firm-destination pairs (namely, the flows that are the base units of our analysis). As a first step, one has to transform the panel of transactions into a matrix of time series, one for each firm-destination pair, and whenever a transaction is not observed for a given time step a zero is imputed. Figure 3 offers a visual representation of this operation. Let us notice that, since our analysis focuses on the intensive margin of export flows, when estimating the model on logarithmic growth rates the imputed zeros generate NAs, which incidence and distribution has to be analyzed in order to proceed with the model estimation.

Dataset sparsity and skewness along the country dimension The first relevant issue arising in the attempt to estimate the model (1) regards the sparsity of the dataset: 89.99% of all elements are missing values with a very skewed distribution across flows. Figure 4 shows that, after the basic cleaning procedures, almost the 90% of the series has more than 75% of missing.

FirmID	Dest.	Value	Year	Quarter
00215	DE	6126	1993	04
00215	DE	8114	1994	02
00215	DE	1134	1994	03
00215	IT	474469	1993	03
00215	IT	127050	1993	04
00215	IT	55780	1994	03
00215	IT	357415	1994	04

Time flow →

FirmID	Dest.	...	1993.03	1993.04	1994.01	1994.02	1994.03	1994.04	...
00215	DE	...	-	6126	-	8114	1134	-	...
00215	IT	...	474469	127050	-	-	55780	357415	...

Figure 3: The transformed series. From the table at the bottom, yearly growth rates are calculated on four points per years on a yearly basis, taking quarter-to-quarter logarithmic ratios.

The second problem concerns the skewness along the country dimension. Among the 259 included in the original dataset, only few of them are relevant for our analysis: for example, during the considered time span only 65 countries report on average at least 1000 active flows (out of potential 3.2 million of flows) and this, of course, reflects into the distribution of export value.

Time frequency, spatial and other restrictions Given the properties of the dataset, the manipulations aimed to design the correct setup to estimate the equations of the model are related, on one side, to the choice of the optimal time frequency and,

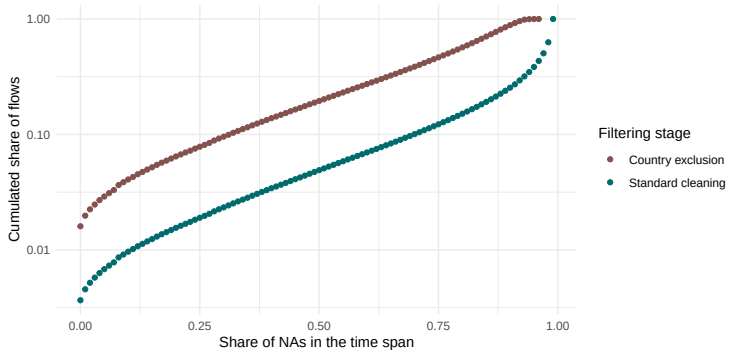


Figure 4: Cumulated share of active flows (y axis in log scale) with increasing shares of missing values in the time span (on a quarterly basis).

on the other side, to the possibility of reducing the number of countries and, hence, the model's factors space dimension. Concerning the frequency, we rely on quarterly data in attempt to enhance the standard volatility analysis in two different characteristics. First, quarterly data allow to deal with sufficiently long time series (96 points in yearly quarter-to-quarter growth rates) providing a proper context for the identification of growth rates dynamics. Second, taking quarters in place of years reduces possible biases due to the so called partial-year effect (see e.g. Bernard et al., 2017), that might lead to the overestimation of the growth rates between the first and the second year (and therefore of the associated volatility) because firms start exporting at different months during the first year of activity.

We next consider how to restrict the number of countries to exclude those that are least interested by trade flows of French firms. In this respect, a robust and consistent estimation of the destination-specific effects requires that factors have measurable impacts both at macroeconomic and microeconomic level. We keep in the dataset those countries that: i) are sufficiently represented in the portfolios of the firms; ii) are relevant in terms of export value as share of the total export. This leaves us with 65 destinations, accounting for 88.25% of total export value. As we are ultimately interested in growth rates, we further drop firm-country flows that over the whole span of time report data on only three points (over 96). We are finally left with 86.44% of total export value and close to 900 thousands firm-destination pairs. Not surprisingly, the operated spatial restriction induces a reduction on micro-level sparsity (see Figure 4).[9]

Growth rates volatility and diversification While the base unit of our statistical analysis will be the firm-destination pair, to provide directly comparable evidence to previous works, let us start reporting descriptive evidence on firm-level volatility and its link with diversification. Defining the weighted logarithmic growth rates of the firms as in (6):

$$y_t^{(e)} = \sum_{d \in I_e} w_{d,t}^{(e)} \, y_{de,t} \quad . \tag{7}$$

[9] After imposing the restrictions, the number of firms in the dataset shrink from around 167 thousands to 140 thousands.

these show the classical tent-shaped distribution confirming the presence of extreme

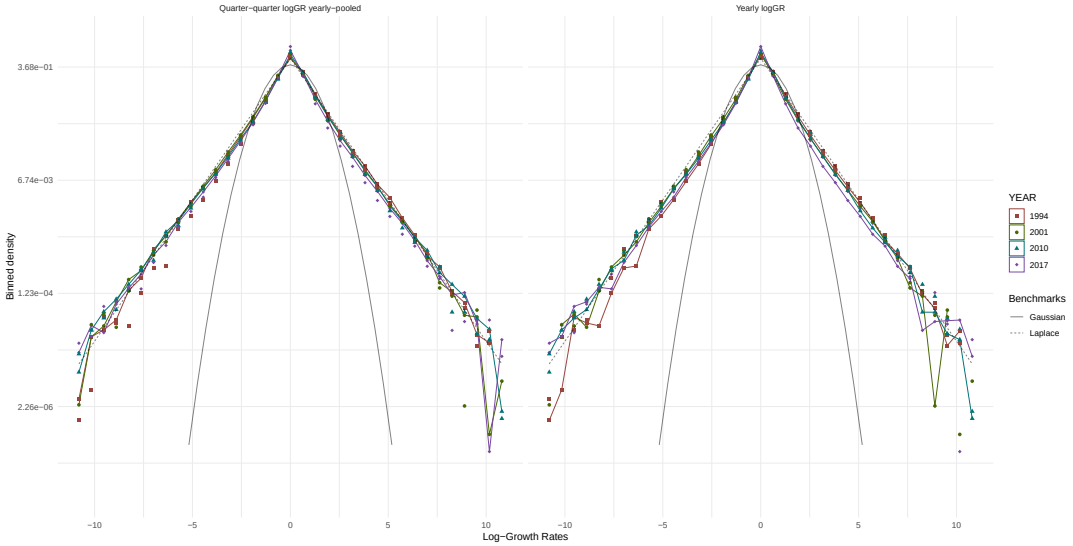


Figure 5: Gaussian kernel densities for firm-level growth rates (as from (7)) at four different moments of the considered time span. Left panel: yearly pooled quarterly GRs. Right panel: yearly GRs.

growth events. As shown in Figure 5, the Laplace density fits properly at different years. The volatility associated to firm s growth rates is computed as its variance along the time dimension. Table 1 reports descriptive statistics for the volatility associated to weighted log-growth rates on a yearly, $\sigma_y(g)$, and a quarterly, $\sigma_q(g)$, basis.

In order to get a first understanding of the relationship between growth rate volatility and patterns of firms' diversifcation, we employ some diversification measures that have been used in the literature: the number (#) of destination markets, the share of firm exports accounted for by the most important market (top share), and the Herfindahl index of export shares (see, among many others, Braakmann and Wagner, 2011; di Giovanni et al., 2014; Vannoorenberghe et al., 2016; Kramarz et al., 2020). For each of these variables, we compute the firm average over time (on a quarterly basis) and then report basic descriptives in Table 1. From the reported statistics, taking as an example the number of destination per year (Figure 6), we observe a consistent level of skewness and concentration in all the relevant distributions.

	N. obs.	Mean	SD	Median	Min	Max	Skewness	Kurtosis
Avg. # of dest.	166928	1.64	4.49	0.31	0.03	134.15	7.71	95.75
Median # of dest.	166928	1.25	4.74	0.00	0.00	134.00	8.12	98.89
Avg. top sh.	166928	0.24	0.19	0.18	0.01	1.00	1.20	0.94
Avg. Herf. index	166928	0.78	0.17	0.83	0.00	0.99	-1.36	1.68

Table 1: Summary statistics for the distributions of firm-level volatility, on a quarterly and yearly basis, and for some key diversification indicators. Averages and median are taken on the distribution of values across the active quarters of the time span of a firm.

Once we have identified a palette of diversification indicators, whose robustness will be discussed in the following, we can give a first glance to the pairwise correlations. Table 2 provides a descriptive confirmation that more diversified firms tend to

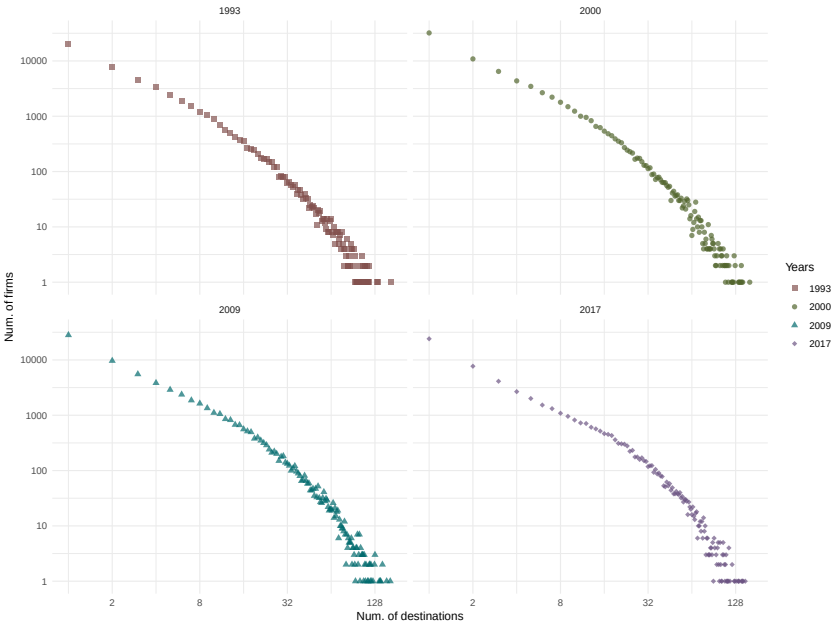


Figure 6: Firms per number of active destinations in log-log scale at different moments of the considered time span.

experience less volatile growth patterns. While discussing the results on the volatility decomposition this relationship will be analyzed further looking at the distributional properties of the volatility (and its components) for classes of firms grouped by diversification quantiles (see the final part of section 4).

	$\sigma(g)$	Avg. #	Med. #	Avg. S_{TOP}	Avg. Herf.
$\sigma_q(g)$	1.0000				
Avg. #	-0.1396	1.0000			
Median #	-0.1185	0.9578	1.0000		
Avg. S_{TOP}	-0.1151	0.1982	0.2033	1.0000	
Avg. Herf.	0.0835	-0.0772	-0.0912	-0.9844	1.0000

Table 2: Correlation matrix of the indicators of Table 2. Different diversification measures (on a quarterly basis when not specified) and log weighted volatility on yearly and quarterly basis.

4 Results

In this section we outline an overview of the main results of our analysis, grouped in two main categories. First, on the macroeconomic dimension, we show how dynamic factor models provide a robust identification of macroeconomic factors that, together with flow-specific loadings, serve as the main tool for the volatility decomposition. For the macroeconomic analysis, the latter is recovered in a comparative framework, where the implications of different aggregation strategies are shown, and then used to measure the relative impact of the global, destination-specific and idiosyncratic components. On the same line, it is shown how absolute volatility estimates distribute along geographical patterns typical of gravity models for export flows. Second, on the

microeconomic dimension, we look at the distributions of the components of firms' volatility and then provide a characterization of the linkages between spatial diversification and volatility trends.

4.1 Volatility at the aggregate level

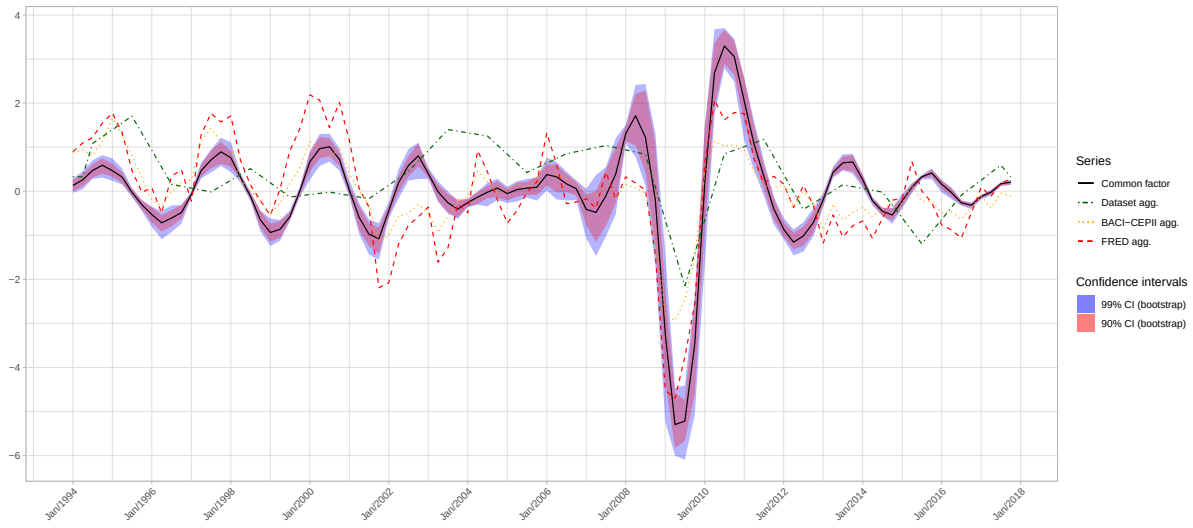


Figure 7: The identified global factor (black full line) with 90% and 99% confidence intervals, compared with other independent sources for aggrgate growth rates of French export: i) FRED (red dashed line), ii) BACI-CEPII (yellow dotted line), iii) the simple aggregate series of the dataset (green dot-dashed line).

Factor space reconstruction The global factor identified is shown in Figure 7, where it is compared with French total export quarterly growth rates from FRED database. Simple visual inspection suggests a certain level of agreement between the two independent measures of total export. This means that the chosen empirical strategy finds and exploits the nexus between micro-level transactions and aggregate macroeconomic trends: comovements of the export flows encode enough information to reconstruct the behaviour of aggregate statistics at a first level of approximation. What clearly remains unexplained by the factor are second order movements and possible shifts of the complete time series: this is natural if we recall that the latent factors have zero mean by definition and come out from a smoothing procedure that cleans second order fluctuations.

Before moving to the the use of this result in the analysis of aggregate volatility, we need a further consistency check on the factor space identified via the EM algorithm, due to the special procedure of sampling applied for the estimation of the model's parameter. To check whether factors' sample estimates describe the same factor space we compute pairwise trace statistics with the formula[10]:

[10] The matrix Z denotes in short both global and destination specific factors.

$$\mathrm{Tr}_{(k,h)} = \frac{\mathrm{Tr}\left(\hat{Z}^{(k)\prime}\hat{Z}^{(h)}\left(\hat{Z}^{(h)}\hat{Z}^{(h)\prime}\right)^{-1}\hat{Z}^{(h)\prime}\hat{Z}^{(k)}\right)}{\mathrm{Tr}\left(\hat{Z}^{(k)\prime}\hat{Z}^{(k)}\right)}$$

The range of the trace statistics is $(0, 1)$ and the factor space tend to be closer when $\mathrm{Tr}_{(k,h)}$ approaches the right limit. Trace statistics allow to test for the equivalence of the factor spaces estimated running the EM algorithm on each sample. The estimated matrices of factors from two different sample are compared taking the related trace statitistics. The pairwise computed values for 20 samples have a minimum of 0.96 and maximum of 0.98, confirming a very good coherence of the different estimates.

Aggregate volatility and granularity The estimated factors and loadings serve as the starting points to proceed to the identification of the macroeconomic components of the volatility. As anticipated, distributional effects might play an effective role in determining the relative importance of the volatility components and results might show consistent variations if the actual shapes of the empirical distributions are taken into account. After a simple visual inspection of the fitted densities (see e.g. the bottom panel of Figure 8 for the global/common flow-level component), we draw attention to two relevant facts: first, we observe Laplace-like shapes for all the years and all the components; second, the evolution of the fitted parameters display remarkably different time-patterns for the three components, the common/global components showing consistent fluctuations during the trade collapse, attributable either to cyclical transfers of mass among the tails or shift to the mean of the distribution. Figure 8 displays the time series of the components' means[11] and the fitted distributions for the global component only at four selected time steps. A detailed characterisation of the evolution of the fittings is out of the scope of this paper, nonetheless this preliminary analysis suggests that, while aggregating, if the distributional movements of the components are encoded the relative importance of the volatility components might change.

Having identified the common shock, we identify the destination specific shocks and the residual idiosyincratic shock as per Equation 1. Using these results we first assess the impact on the three sources of shocks on the volatility of aggregate export sales. The left panel of Figure 9 reports the relative volatility, measured as the ratio between the standard deviation explained by the single components and the overall variance of the growth rates, [12], whereas the right panel shows the time-correlation with export sales' time series. It is evident that variation in aggregate export is dominated by the idiosyncratic component, rather than global and destination-specific ones. The volatility associated to firm level shocks is close to 0.8 of the total volatility and presents a high level of correlation. By contrast, global and destination-specific contributions are relatively less variable, with a clear prominence of the former on the latter, yet still with a positive and significant correlation. These results confirm in part the general knowledge of the dominance of idiosyncratic shocks in determining aggregate

[11] E.g. the series $\mu_t^{\mathrm{glob}} = \frac{1}{\#(I)}\sum_{i\in I}\hat{\lambda}_i\hat{f}_t$ for the global/common component

[12] In the context of a decomposition, we name them 'shares', notice however that due to the existence of covariance terms, they do not sum to one.

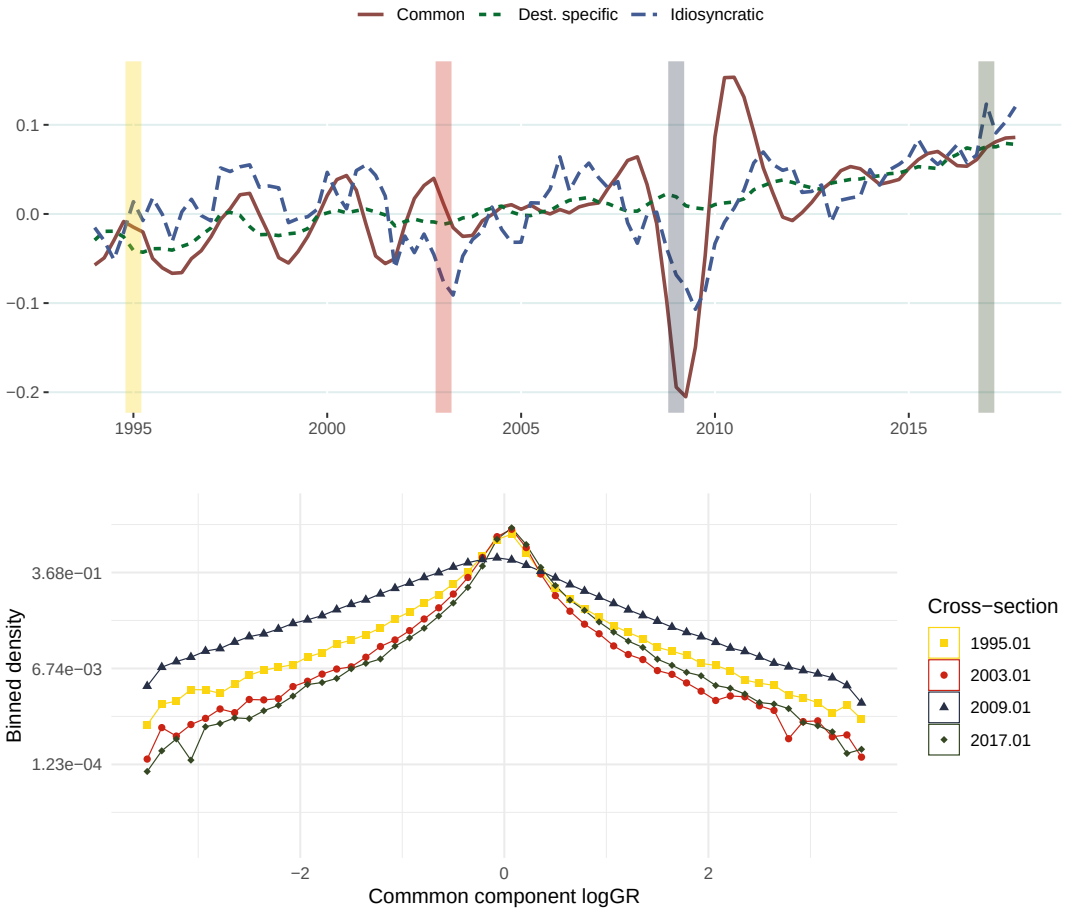


Figure 8: Zooming into the cross-sectional distributions of the growth rates components. In the top panel, the simple component means at each time step form the time series from 1993 to 2017. On the bottom panel, the distribution of the common component is fitted at four selected time-points, showing considerable differences in the distribution of probability mass in the tail and at the center of the distribution. This suggests that while cross-section averages can partly capture this movements, considering higher order moments in the aggregation procedure might better capture the underlying dynamics.

volatility (Gabaix, 2011; Carvalho and Gabaix, 2013; di Giovanni et al., 2014) with a major distinction regarding the measure of this impact. Indeed, in measuring global and destination-specific impacts as the interplay between a latent common factor and flow-specific coefficients, the model captures more variation, possibly because the two components encode both common movements and the implied idiosyncratic responses. This for example might explain why the non idiosyncratic component explains around 30% of the aggregate export volatility, which seems to be larger than the share accounted for by the non-idiosyncratic component in the analysis by di Giovanni et al. (2014). With respect to the latter, another possible complementary explanation could be found in the peculiar time span of our analysis: the inclusion of the trade collapse (Baldwin, 2009) and the subsequent recovery gives more relevance to common movements implied by the synchronization typical of deep downturns and following rebounds[13]. As anticipated, changing aggregation strategy increases the relative im-

[13] On this, in Figure 7 the almost perfect match between the factor and the aggregate suggests that

portance of the global/common components.

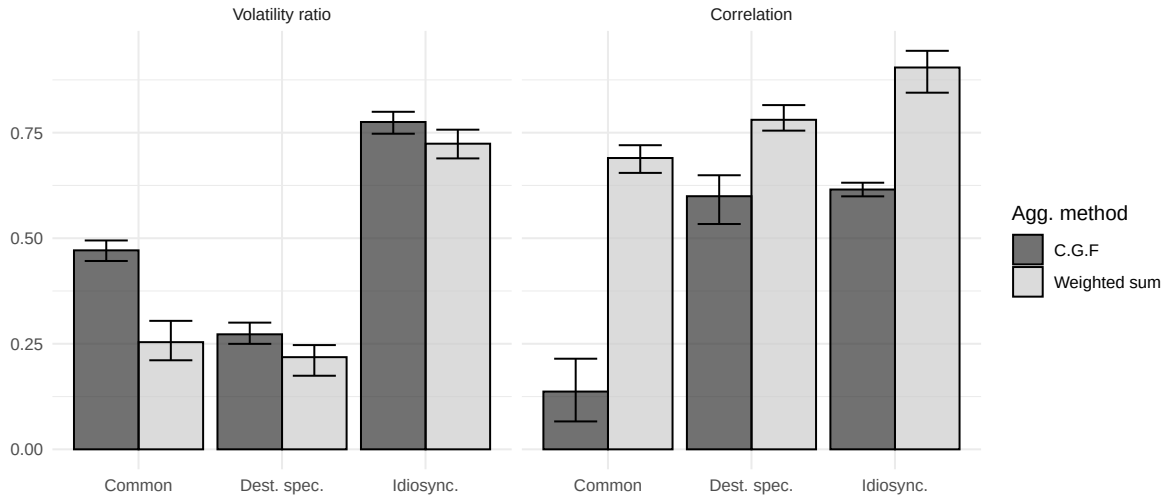


Figure 9: Aggregate volatility decomposition. Correlation and volatility share of the three components as simple averages over the 20 sample estimations (each samples counting 80k firms). Error bars define the confidence interval over the repeated sample estimations. At the bottom of each bar: the ratio between the measure of the 90% confidence interval and the estimated value of the related statistics.

Macroeconomic geographical patterns. Looking at the destination-specific volatility associated to the aggregate French export, we are able to identify gravity-like patterns confirming that geographic and spatial effects — affecting flows and trade intensities — mitigate the risks attached to high volatile trade relationships. We measure the volatility associated to a given export destination as

$$\mathrm{Var}_T\left(\sum_{i\in I_c}\omega^c_{i,t}(\lambda_i\cdot f_t+\rho_{c,i}\cdot d_{c,t})\right) \tag{8}$$

where I_c is the set of the flows targeting country c and $w^c_{i,t}$ is the relative dynamic weight associated to the flow $i \in I_c$. A preliminary inspection of the scatter plots in Figure 10 confirms the clear inverse relation when restricting the analysis to EU countries, whereby the EU single market provides a common playground for trade relationship mitigating distortions coming from different economic factors. Indeed, the outliers, if any, are considered intra EU commercial partners joining the Union in the second half of the time window[14]. Analogous considerations apply to the patterns on the left panel: whenever a sharp inverse relation is violated it is possible to identify underlying macroeconomic mechanisms fostering or dampening both number and intensity of microeconomic trade flows. Among the many, we highlighted the trade relationship with ex-colonies, reducing the risks expected for commercial partnerships

synchronization is influencing the outcome.

[14] Slovakia and Lithuania joined the EU single market in 2004 whether the considered time span goes from 1993 and 2017

with low income countries as shown by the downward deviation of the relative points. To provide a quantitative assessment of these trends, we ran a simple gravity-like regression adding geographical distance and controls grounded on geopolitical variables (free trade agreements, ex-colonies relationship, intra or extra EU dummies). Results presented in Table 3 and Figure 10 confirm the evidence and provide a solid quantitative ground to the outlined macroeconomic trends.

Figure 10: Destination-specific volatility vs destination GDP for extra-EU (left panel) and intra-EU (right panel) trade relations.

	Log. dest. vol.		
	(1)	(2)	(3)
Log. dist.	0.251***	0.243***	0.195***
	(0.064)	(0.079)	(0.072)
Log. GDP	0.009	−0.106**	−0.084**
	(0.031)	(0.040)	(0.040)
Geo+Colonies controls	No	Yes	No
FTA+Colonies controls	No	No	Yes
Constant	−3.759***	−0.505	−0.628
	(0.995)	(1.303)	(1.138)
Observations	65	65	65
R^2	0.199	0.378	0.431
Adjusted R^2	0.173	0.326	0.350
Residual Std. Error	0.488 (df = 62)	0.441 (df = 59)	0.433 (df = 56)
F Statistic	7.680*** (df = 2; 62)	7.179*** (df = 5; 59)	5.298*** (df = 8; 56)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3: OLS estimates of gravity-like regressions for the volatility associated to each destination.

4.2 Volatility at the firm level

In order to establish the role of the various sources of shocks to growth rates fluctuations (and hence determining volatility), in line with the work of Kramarz et al. (2020), we look at micro-distributions where one or more components have been muted. Comparing the two bottom panels of Figure 11 we see that global and destination-specific components have similar impacts on firm volatility distribution, both in terms of magnitude and direction: once muted singularly, a visible effect consists in a shift to the left of the second and third quartile threshold and, with minor intensity, of the first one; the effect is almost identical for both components. On the other hand, if we exclude the idiosyncratic component we observe a relevant left-shift of the quartile thresholds and a substantial narrowing of the right tail. More precisely, the median reduction due to the removal of one of the aggregate shocks is around the 30% (from 0.93 to 0.65), whereas the impact of the microeconomic contribution amounts to a dampening of the 47% (from 0.19 to 0.8), confirming the prominent role of the idiosyncratic component, yet showing global and destination-specific terms have a non negligible effects as drivers of volatility. This means that exporters, even though majorly exposed to microeconomic risks, still face the risk of shocks coming from aggregate fluctuations lead by the global and regional macroeconomic factors. In the remainder, the focus will move to the nature of this link and the role of diversification in mitigating its effects.

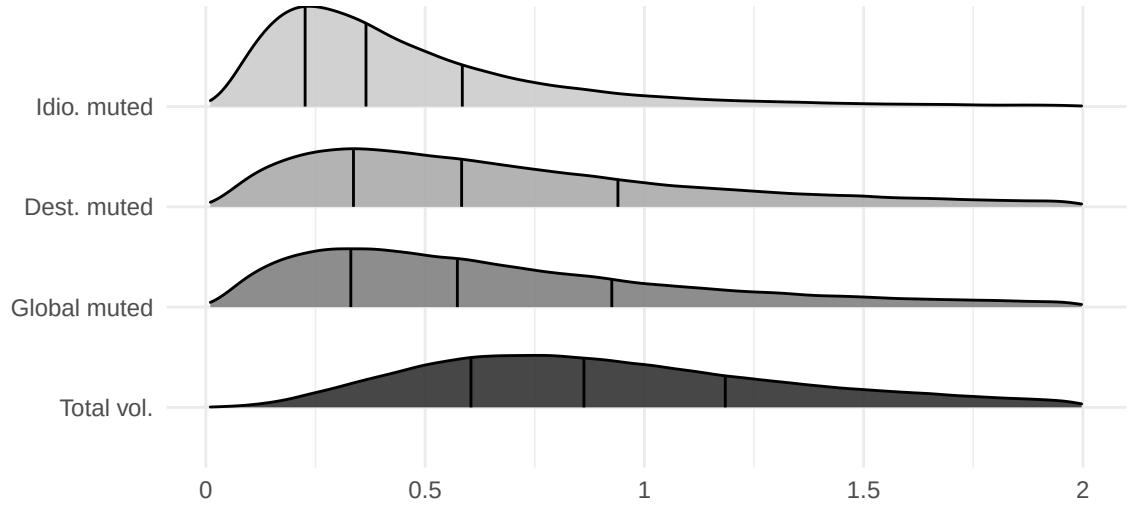


Figure 11: Gaussian kernel densities for 'counterfactual' volatility distributions: on each panel a fit of the empirical distribution is provided for firm-level volatilities silencing the macroeconomic effects together (top panel) and singularly (middle panels). A comparison is offered with the actual volatility distribution (bottom panel).

Volatility components and diversification The preceding analysis has two main implications. First, that firm-level volatility has a fundamental idiosyncratic component, which by definition is not affected by global and destination specific shocks.

Second, that, on top of a major firm-specific component, macroeconomic business cycles play a non negligible role as drivers of volatile growth paths. At this stage, one might wonder whether diversification strategies help firms to reduce trade-related risks, and how and to what extent they impact the different components. Common wisdom suggests that we should observe a negative relationship between firm level volatility and degree of diversification in terms of destination markets. Table 4 provides a quantitative assessment of a linear relation between the logarithm of firm volatility and diversification — measured in terms of inverse Herfindahl index, as time average across the considered window — showing that the linear fit is enhanced when controlling for sectoral patterns and size classes.

An analogous exercise can be repeated for the three different components. When fitting a linear relation (Table 5), diversification seems to help dampening the risks coming from macroeconomic sources more than microeconomic ones, confirming the idea that flow-specific idiosyncratic risks can be hardly diversified away. Actually, since the exporters' universe is populated by highly heterogeneous players and diversification strategies have non-linear effects, we present a deeper exploration beyond the realm of linear regressions. In fact, the low R-squared estimates suggest that the interplay between volatility and diversification develops on more complex patterns: moving along the quantiles of diversification index and of estimated volatility, we observe contrasting trends. A visual inspection of the quantile-quantile plots might help understanding the underlying dynamics (Figure 12 for general trends and Figure 13 high and low volatile firms). First, the macroeconomic volatility components move together — if we consider both the most and the least volatile firms — with a downward trend in logs. Second, the idiosyncratic component moves along two opposite trends if we look at exporters that diversify respectively below or above the median. Indeed, looking at the first half of the diversification spectrum the volatility lies on a steady path, or even increases for less volatile firms, whereas on the second half moves along the expected inverse linear path, meaning that risk mitigation becomes relevant only after a certain threshold. This diversification limit is indeed quite high, corresponding to an average Herfindahl index of 0.10, that is the theoretical values obtained by a firm operating evenly on ten different destinations for her life span.

Summing up, the risk exposure is reduced for firms that diversifies their activities on the destination markets. Log-linear risk dampening effects seem to work only for shocks originated by macroeconomic induced fluctuation, with no remarkable difference between global and destination-specific shocks. On the contrary, the idiosyncratic component of the growth rate generates a volatility distribution at the firm level that does not change substantially while firms diversify more until a certain level, when passing the threshold diversification strategies give a consistent reduction helping firms to approach less volatile growth paths.

	Dependent variable:			
	log(VOL)			
	(1)	(2)	(3)	(4)
Div. index	0.074^{***}	-0.023^{***}	-0.182^{***}	-0.195^{***}
	(0.004)	(0.004)	(0.011)	(0.010)
Log Size (2nd decile)		0.837^{***}	0.582^{***}	0.487^{***}
		(0.023)	(0.095)	(0.092)
Log Size (3rd decile)		1.314^{***}	0.906^{***}	0.827^{***}
		(0.023)	(0.095)	(0.091)
Log Size (4th decile)		1.606^{***}	1.128^{***}	1.036^{***}
		(0.023)	(0.096)	(0.092)
Log Size (5th decile)		1.842^{***}	1.325^{***}	1.163^{***}
		(0.023)	(0.095)	(0.091)
Log Size (6th decile)		1.953^{***}	1.455^{***}	1.294^{***}
		(0.023)	(0.096)	(0.092)
Log Size (7th decile)		2.051^{***}	1.497^{***}	1.344^{***}
		(0.023)	(0.094)	(0.090)
Log Size (8th decile)		2.070^{***}	1.427^{***}	1.255^{***}
		(0.023)	(0.095)	(0.091)
Log Size (9th decile)		1.994^{***}	1.369^{***}	1.246^{***}
		(0.023)	(0.092)	(0.088)
Log Size (10th decile)		1.694^{***}	0.677^{***}	0.644^{***}
		(0.023)	(0.091)	(0.088)
Sector control				YES
Log Size Q. x Div. Q			0.557^{**}	0.666^{***}
Constant	-1.545^{***}	-2.891^{***}	-1.740^{***}	-2.295^{***}
	(0.009)	(0.017)	(0.072)	(0.083)
Observations	143,194	140,310	140,310	140,310
R^2	0.002	0.098	0.182	0.246
Adjusted R^2	0.002	0.098	0.181	0.245
Residual Std. Error	1.999 (df = 143192)	1.889 (df = 140299)	1.800 (df = 140199)	1.729 (df = 140103)
F Statistic	330.568^{***} (df = 1; 143192)	$1{,}530.720^{***}$ (df = 10; 140299)	283.482^{***} (df = 110; 140199)	221.440^{***} (df = 206; 140103)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 4: The results for the OLS estimates of the regression of the logarithm of the firm-level volatility against the inverse of the Herfindahl–Hirschman Index (HHI), with four incremental model specification: i) plain regression, ii) controlling for firms' size class (deciles), iii) including the interaction term between diversification class (percentiles) and size class (deciles), iv) controlling for sector fixed-effects.

	Dependent variable:		
	log(COM_VOL)	log(CNT_VOL)	log(IDIO_VOL)
	(1)	(2)	(3)
Div. index.	-0.269^{***}	-0.244^{***}	-0.190^{***}
	(0.010)	(0.010)	(0.010)
Log Size (2nd decile)	0.207^{***}	0.232^{***}	0.678^{***}
	(0.026)	(0.024)	(0.026)
Log Size (3rd decile)	0.357^{***}	0.378^{***}	1.131^{***}
	(0.027)	(0.025)	(0.027)
Log Size (4th decile)	0.395^{***}	0.410^{***}	1.437^{***}
	(0.028)	(0.026)	(0.028)
Log Size (5th decile)	0.481^{***}	0.497^{***}	1.722^{***}
	(0.028)	(0.027)	(0.029)
Log Size (6th decile)	0.479^{***}	0.524^{***}	1.892^{***}
	(0.029)	(0.028)	(0.030)
Log Size (7th decile)	0.553^{***}	0.531^{***}	2.096^{***}
	(0.030)	(0.028)	(0.031)
Log Size (8th decile)	0.438^{***}	0.517^{***}	2.152^{***}
	(0.030)	(0.029)	(0.031)
Log Size (9th decile)	0.453^{***}	0.454^{***}	2.249^{***}
	(0.031)	(0.029)	(0.032)
Log Size (10th decile)	0.208^{***}	0.295^{***}	2.094^{***}
	(0.031)	(0.030)	(0.032)
Sector control	YES	YES	YES
Size Q. x Div. Q.	1.288^{***}	1.209^{***}	1.564^{***}
	(0.068)	(0.065)	(0.070)
Constant	-3.855^{***}	-4.055^{***}	-4.178^{***}
	(0.049)	(0.046)	(0.050)
Observations	140,310	140,310	140,310
R^2	0.131	0.137	0.258
Adjusted R^2	0.130	0.135	0.257
Residual Std. Error (df = 140103)	1.678	1.593	1.722
F Statistic (df = 206; 140103)	102.629^{***}	107.608^{***}	236.396^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5: Results of the OLS estimates for the regressions of the three components against the diversification index with size and sector controls (the Herfindahl–Hirschman Index).

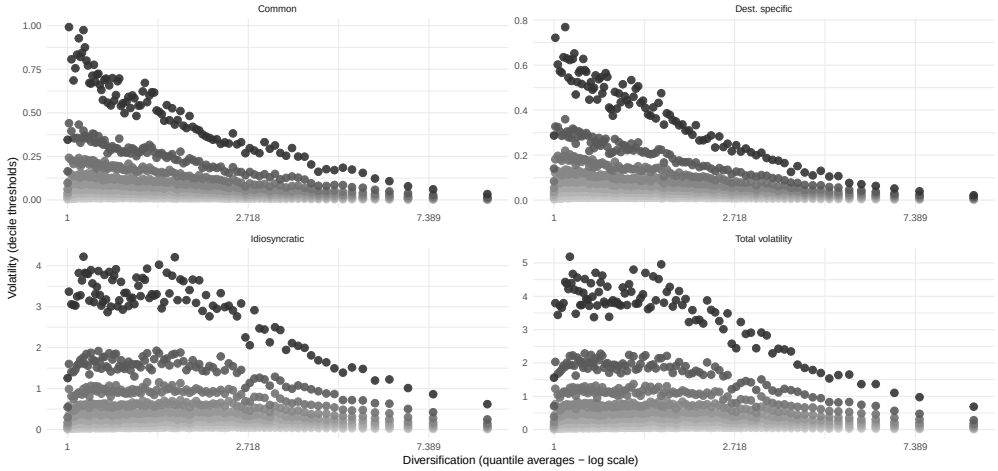


Figure 12: Diversification versus volatility, quantile-quantile plot. On the y-axis the deciles of volatility are plotted, from the darkest (top 10%) to the lightest (bottom 10%).

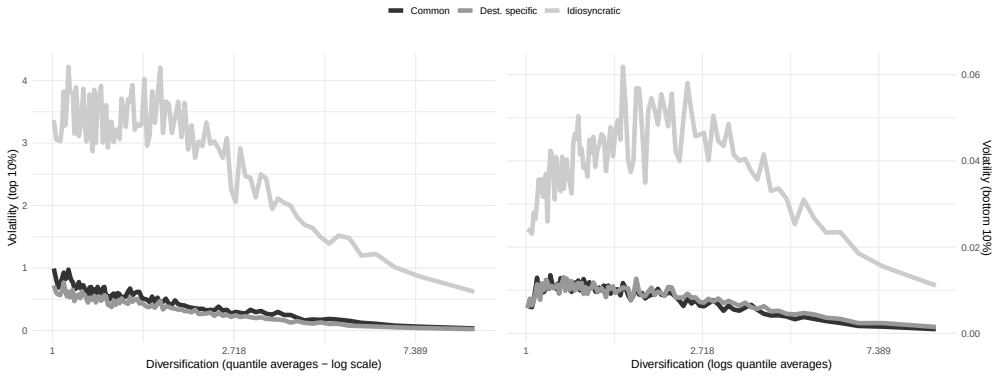


Figure 13: Diversification versus volatility, quantile-quantile plot. On the left panel the first decile of the volatility distribution for different percentiles of inverse Herfindahl index (average on the time window), for the three different components. On the right panel, the analogous graph for the bottom 10%.

5 Conclusions

In this paper we propose a dynamic factor model approach to the decomposition of aggregate and firm-level volatility. Aspects of novelties are related to the estimation technique adopted to reconstruct the latent space of macroeconomic and destination-specific shocks. The econometric strategy provides manifold improvements to the extant literature. First, we address the decomposition problem in a proper dynamic framework proposing an orthogonal partition whose terms account for the evolution along the time dimension of the macroeconomic factors (described as simple AR processes). Second, we disentangle the sources of macroeconomic fluctuations separating flow-specific response to global and to destination-specific shocks. This provides new insights on the role of the global/common component to the volatility, which role has always been considered minor with respect to the idiosyncratic component. Third, at the microeconomic level we are capable to analyze the relevant distributions of the various component, showing that the contributions of the global and destination-specific

effects are comparable in the way they affect the first and the second moment of the distribution, whereas the idiosyncratic component of the volatility has a much more sizeable effect. In addition, we show how diversification across destination markets can protect firms from shocks coming from macroeconomic events but seems not to have effect on the idiosyncratic components. Concluding, not to be overlooked is the methodological contribution. To our knowledge, we provide the first application of dynamic factor model applied to microeconomic data, which required an estimation based on a modified expectation maximization algorithm to accommodate both the missing values and the high numerosity of the micro time series, opening the way for other similar applications to firm-level analysis.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012): "The network origins of aggregate fluctuations," *Econometrica*, 80, 1977–2016.

Amiti, M. and D. E. Weinstein (2018): "How much do idiosyncratic bank shocks affect investment? Evidence from matched bank-firm loan data," *Journal of Political Economy*, 126, 525–587.

Baldwin, R. E. (2009): *The great trade collapse: Causes, consequences and prospects*, Cepr.

Bańbura, M. and M. Modugno (2014): "Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data," *Journal of Applied Econometrics*, 29, 133–160.

Barigozzi, M. and M. Luciani (2019): "Quasi maximum likelihood estimation and inference of large approximate dynamic factor models via the EM algorithm," *arXiv preprint arXiv:1910.03821*.

Bergounhon, F., C. Lenoir, and I. Mejean (2018): "A guideline to French firm-level trade data," Tech. rep., mimeo Polytechnique.

Bernard, A. B., E. A. Boler, R. Massari, J.-D. Reyes, and D. Taglioni (2017): "Exporter dynamics and partial-year effects," *American Economic Review*, 107, 3211–28.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2009): "The Margins of US Trade," *American Economic Review*, 99, 487–93.

——— (2012): "The Empirics of Firm Heterogeneity and International Trade," *Annual Review of Economics*, 4, 283–313.

——— (2016): "Global Firms," CEP Discussion Papers dp1420, Centre for Economic Performance, LSE.

Bottazzi, G., L. Li, and A. Secchi (2019): "Aggregate fluctuations and the distribution of firm growth rates," *Industrial and Corporate Change*, 28, 635–656.

Bottazzi, G. and A. Secchi (2006): "Gibrat's law and diversification," *Industrial and Corporate Change*, 15, 847–875.

Braakmann, N. and J. Wagner (2011): "Product diversification and stability of employment and sales: first evidence from German manufacturing firms," *Applied Economics*, 43, 3977–3985.

Breitung, J. and S. Eickmeier (2015): "Analyzing business cycle asymmetries in a multi-level factor model," *Economics Letters*, 127, 31–34.

Carvalho, V. and X. Gabaix (2013): "The Great Diversification and Its Undoing," *American Economic Review*, 103, 1697–1727.

Carvalho, V. M. and B. Grassi (2019): "Large firm dynamics and the business cycle," *American Economic Review*, 109, 1375–1425.

Caselli, F., M. Koren, M. Lisicky, and S. Tenreyro (2020): "Diversification Through Trade*," *The Quarterly Journal of Economics*, 135, 449–502.

di Giovanni, J. and A. A. Levchenko (2009): "Trade Openness and Volatility," *The Review of Economics and Statistics*, 91, 558–585.

di Giovanni, J., A. A. Levchenko, and I. Mejean (2014): "Firms, Destinations, and Aggregate Fluctuations," *Econometrica*, 82, 1303–1340.

Di Giovanni, J., A. A. Levchenko, and I. Mejean (2018): "The micro origins of international business-cycle comovement," *American Economic Review*, 108, 82–108.

Doz, C., D. Giannone, and L. Reichlin (2012): "A quasi–maximum likelihood approach for large, approximate dynamic factor models," *Review of economics and statistics*, 94, 1014–1024.

Durbin, J. and S. J. Koopman (2012): *Time series analysis by state space methods*, Oxford university press.

Eaton, J., S. Kortum, and F. Kramarz (2004): "Dissecting Trade: Firms, Industries, and Export Destinations," *American Economic Review*, 94, 150–154.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The generalized dynamic-factor model: Identification and estimation," *Review of Economics and statistics*, 82, 540–554.

Gabaix, X. (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79, 733–772.

Kelly, B., H. Lustig, and S. V. Nieuwerburgh (forthcoming): "Firm Volatility in Granular Networks," *Journal of Political Economy*.

Koren, M. and S. Tenreyro (2007): "Volatility and Development," *The Quarterly Journal of Economics*, 122, 243–287.

Kramarz, F., J. Martin, and I. Mejean (2020): "Volatility in the small and in the large: The lack of diversification in international trade," *Journal of International Economics*, 122, 103276.

Long Jr, J. B. and C. I. Plosser (1983): "Real business cycles," *Journal of political Economy*, 91, 39–69.

Reis, R. and M. W. Watson (2010): "Relative goods' prices, pure inflation, and the Phillips correlation," *American Economic Journal: Macroeconomics*, 2, 128–57.

Stock, J. H. and M. W. Watson (2002): "Forecasting using principal components from a large number of predictors," *Journal of the American statistical association*, 97, 1167–1179.

Vannoorenberghe, G., Z. Wang, and Z. Yu (2016): "Volatility and diversification of exports: Firm-level theory and evidence," *European Economic Review*, 89, 216 – 247.

A Estimation strategy: technical details

A.1 Dynamic factor models with arbitrary pattern of missing data in presence of a block structure

The estimation of the main equation of the model with additional dynamic equations on the common factors formally pass through the definition of a dynamic factor model with an imposed block structure. We propose here a general version of the model with a number of blocks C, an arbitrary number of common factor (K_G) and an arbitrary number of factors per block (K_c with $c = 1, \ldots, C$)[15].

$$y_{i,t} = \lambda_i f_t + \rho_{c,i} d_{c,t} + \xi_{i,t} \qquad \xi_t \sim \text{IID}(0, R) \tag{9}$$

$$f_{g,t} = a_g f_{g,t-1} + u_{g,t} \qquad u_{g,t} \sim \text{IID}(0, q_f) \tag{10}$$

$$d_{l_c,t} = a_c d_{l_c,t-1} + u_{l_c,t} \qquad u_{l_c,t} \sim \text{IID}(0, q_c) \tag{11}$$

where the index $i = 1, \ldots, N$ runs over the export flows, $t = 1, \ldots, T$ over the time span , $c = 1, \ldots, C$ over the selected blocks (countries), $l_c = 1, \ldots, K_c$ over the country-specific number of factors, $g = 1, \ldots, G$ over the global factors and ξ_t defines the vector of the idiosyncratic terms at the cross-section. As for the elements of the equation, $f_{g,t}$ is a global factor whose dynamics is ruled by (10) and $d_{l_c,t}$ are K_c factor specific to the block c. Each time-series is associated to only one block, hence we have that the number of flows admits the partition $N = \sum_{c=1,\ldots,C} n_c$, where n_c is the number of time series associated to block c. The system can be written in the synthetic vectorized form:

$$y_t = \Lambda z_t + \xi_t \qquad \xi_t \sim \text{IID}(0, R) \tag{12}$$

$$z_t = A z_{t-1} + \mathbf{u}_t \qquad u_t \sim \text{IID}(0, Q) \tag{13}$$

[15] The estimation of the model (1) is obtained taking $K_G = K_c = 1 \ \forall c$ and as C the number of destination-specific effects.

and, for the case $K_G = K_C = 1$, the block-structure with zero imposed restrictions becomes:

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ \vdots \\ \vdots \\ Y_{C1} \\ \vdots \\ Y_{Cn_R} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \rho_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ \lambda_{1n_1} & \rho_{1n_1} & 0 & \cdots & 0 \\ \lambda_{21} & 0 & \rho_{21} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ \lambda_{2n_2} & 0 & \rho_{2n_2} & \cdots & 0 \\ \vdots & & & \vdots & \\ \vdots & & & \vdots & \\ \lambda_{C1} & 0 & 0 & \cdots & \rho_{C1} \\ \vdots & & \vdots & \ddots & \vdots \\ \lambda_{Cn_R} & 0 & 0 & \cdots & \rho_{Cn_R} \end{pmatrix} \cdot \begin{pmatrix} f_t \\ d_{1,t} \\ \vdots \\ d_{C,t} \end{pmatrix} + \begin{pmatrix} \xi_{11} \\ \vdots \\ \xi_{1n_1} \\ \xi_{21} \\ \vdots \\ \xi_{2n_2} \\ \vdots \\ \vdots \\ \xi_{R1} \\ \vdots \\ \xi_{Rn_R} \end{pmatrix}$$

where F is a $(T \times 1)$ global factor vector and D_c are block factor vectors. Following Doz et al. (2012) we compute the log-likelihood, $l(Y, Z, \theta)$, associated to the system composed by the equations (12) and (13) as a function of the data matrix Y, the set of parameters $\theta = \{\Lambda, R, A, Q\}$ and the matrix of factors Z(see appendix B of Bańbura and Modugno, 2014, for the explicit form.) Then, we proceed to the maximization of l through the following two steps procedure (which details are exposed below):

1. Given the factors, the parameters are derived by analytical maximization of l.

2. Given the parameters, the factors are derived running Kalman Filter/Smoother on the system of equations (12) and (13).

This procedure defines a sequence of increasing log-likelihood values

$$l\left(Y, F^{(0)}, \theta^{(0)}\right) \rightarrow l\left(Y, F^{(0)}, \theta^{(1)}\right) \rightarrow l\left(Y, F^{(1)}, \theta^{(1)}\right)$$

that needs an appropriate initialization and stops when an appropriate stopping rule (we mutuate convergence condition from the end of section 2.1 of Bańbura and Modugno, 2014, with a threshold of 0.001).

Initialization algorithm. The procedure is initialized computing the sequential least square estimator associated to the model on a 'complete' matrix of data. The estimates $\Lambda^{(0)}, R^{(0)}, A^{(0)}, Q^{(0)}$ are the results of the following two steps:

1. We fill the missing values of the original dataset with series medians, then we smooth the outcome taking the moving averages of the series, so that we can work with the filled matrix $\overline{Y}$;

2. Once a complete matrix is given, the sequential least square estimator by Breitung and Eickmeier (2015) can be applied to obtain the block-by-block parameters initialization.

From model parameters to factors: Kalman Filter algorithm. At iteration k, we set up a recursive procedure on the time dimension $t = 1, \ldots, T$ computing (after suitable choices for z_0 and $\widetilde{P}_0$) the predicted state vector and variance-covariance matrix as proposed in Durbin and Koopman (2012). From now on, in order to deal with missing values, parameters are restricted to each time step to that portion with available information. Hence, the "NA" index or suffix denotes the matrix/vector cleaned by row, column or elements corresponding to NA entries at time t. Then, we have:

$$
\begin{aligned}
z_t &= A^{(k)} z_{t-1} && (14)\\
P_t &= A^{(k)} \widetilde{P}_{t-1} A^{(k)\prime} + Q^{(k)} \quad \text{and} \quad G_t^{(k)} = \Lambda_{\text{NA},t}^{(k)} \widetilde{P}_t \Lambda_{\text{NA},t}^{(k)\ \prime} + R_{\text{NA},t}^{(k)} && (15)
\end{aligned}
$$

and the filtered equivalents

$$
\begin{aligned}
\widetilde{z}_t &= z_t + P_t \Lambda_{\text{NA},t}^{(k)\ \prime} G_t^{-1} \left(y_t^{\text{NA}} - \Lambda_{\text{NA},t}^{(k)} z_t \right) && (16)\\
\widetilde{P}_t &= P_t + P_t \Lambda_{\text{NA},t}^{(k)\ \prime} G_t^{-1} \Lambda_{\text{NA},t}^{(k)} P_t' && (17)
\end{aligned}
$$

The varying part of the loglikelihood can be computed as

$$
\begin{aligned}
\text{loglik}^{(k)} = &- \frac{1}{2} \sum_{t=1}^{T} \log \left(\Lambda_{\text{NA},t}^{(k)} \widetilde{P}_t \Lambda_{\text{NA},t}^{(k)\ \prime} + R_{\text{NA},t}^{(k)} \right) && (18)\\
&- \frac{1}{2} \sum_{t=1}^{T} \left(y_t^{\text{NA}} - \Lambda_{\text{NA},t}^{(k)} z_t \right) \left(\Lambda^{(k)} \widetilde{P}_t \Lambda_{\text{NA},t}^{(k)\ \prime} + R_{\text{NA},t}^{(k)} \right) \left(y_t^{\text{NA}} - \Lambda_{\text{NA},t}^{(k)} z_t \right)' && (19)
\end{aligned}
$$

With the filtered and smoothed estimates of z_t and P_t we can initialize the smoother to obtain the smoothed estimates $(\overline{z}_1, \ldots, \overline{z}_T)$, $\overline{Z}$ in matrix form, and $(\overline{P}_1, \ldots, \overline{P}_T)$ by the inverse recursion:

$$
\begin{aligned}
\overline{z}_{t-1} &= \widetilde{z}_{t-1} + \widetilde{P}_{t-1} A' P_t^{-1} \left(\overline{z}_t - A^{(k)} z_{t-1} \right) && (20)\\
\overline{P}_{t-1} &= \widetilde{P}_{t-1} + \widetilde{P}_{t-1} A' P_t^{-1} \left(\overline{P}_t - \overline{P}_{-1,t} \right) \left(\widetilde{P}_{t-1} A' P_t^{-1} \right)' && (21)
\end{aligned}
$$

From latent factors to model parameters. Once the smoothed estimates of the latent factors have been obtained, we can apply the results of the expected log-likelihood maximization over the parameter space, given the observations Ω_T and the parameters $\theta^{(k)}$:

$$
\begin{aligned}
A_g^{(k+1)} &= \left(\sum_{t=1}^{T} E_{\theta(k)} \left[z_{g,t} z_{g,t-1}' | \Omega_T \right] \right) \left(\sum_{t=1}^{T} E_{\theta(k)} \left[z_{g,t-1} z_{g,t-1}{}' | \Omega_T \right] \right)^{-1} && (22)\\
Q_g^{(k+1)} &= \frac{1}{T} \left(\sum_{t=1}^{T} E_{\theta(k)} \left[z_{g,t} z_{g,t}' | \Omega_T \right] \right) - \frac{A_g^{(k+1)}}{T} \left(\sum_{t=1}^{T} E_{\theta(k)} \left[z_{g,t} z_{g,t-1}{}' | \Omega_T \right] \right)' && (23)
\end{aligned}
$$

where the index g denotes the restriction to the global factor. Now, for a given block c, an analogous formula is used for $A_c^{(k+1)}, Q_c^{(k+1)}$. For the actual computations,

expectations used in equations (22) and (23) are replaced by the smoothed estimates of the factors $\overline{Z}$ and their restriction to the common an block specific components, $\overline{Z}_g, \{\overline{P}_g\}_{t=1}^T$ and $\overline{Z}_c, \{\overline{P}_{c,t}\}_{t=1}^T$[16].

$$\hat{A}_g^{(k+1)} = \left(\overline{Z}_g\overline{Z}'_{-1,g} + \sum_{t=1}^{T}\overline{P}_{-1,g,t}\right)\left(\overline{Z}_{-1,g}\overline{Z}'_{-1,g} + \sum_{t=1}^{T}\overline{P}_{g,t}\right)^{-1} \tag{24}$$

$$\hat{Q}_g^{(k+1)} = \frac{1}{T}\left(\overline{Z}_g\overline{Z}'_g + \sum_{t=1}^{T}\overline{P}_{g,t}\right) - \frac{\hat{A}_g^{(k+1)}}{T}\left(\overline{Z}_g\overline{Z}'_{-1,g} + \sum_{t=1}^{T}\overline{P}_{-1,g,t}\right)' \tag{25}$$

$$\tag{26}$$

Estimates of the complete form parameters $\hat{A}^{(k+1)}, \hat{Q}^{(k+1)}$ are obtained by trivial re-composition of the block structure. The updated estimates for the loadings matrix can be seen as the NA-corrected OLS solutions of the regressions (see Bańbura and Modugno, 2014, pag. 138, eq. (11))[17]

$$y_t = \Lambda_g\overline{Z}_{g,t} + \Lambda_c\overline{Z}_{c,t} + v_t \qquad c = 1, \ldots, C \tag{27}$$

for n_b series of the block b.

$$\mathrm{vec}(\Lambda_b) = \left(\sum_{t=1}^{T}\overline{Z}_{gc,t}\overline{Z}'_{gc,t} \otimes \mathrm{Ind}_t^{\mathrm{NA}}\right)^{-1} \mathrm{vec}\left(\sum_{t=1}^{T} y_{b,t}^{\mathrm{NA}}\overline{Z}'_t\right) \tag{28}$$

[18] The matrix $\hat{C}^{(k+1)}$ is composed via vertical merging of the estimated blocks and this step concludes with the matrix R in line with the eq.(12), pag.138 of Bańbura and Modugno (2014):

$$\frac{1}{T}\sum_{t=1}^{T} \left(y_t^{\mathrm{NA}} - \mathrm{Ind}_t^{\mathrm{NA}}\hat{\Lambda}^{(k+1)}\,\overline{Z}_t\right)\left(y_t^{\mathrm{NA}} - \mathrm{Ind}_t^{\mathrm{NA}}\hat{\Lambda}^{(k+1)}\,\overline{Z}_t\right)' + \mathrm{Ind}_t^{\mathrm{NA}}\,\hat{C}^{(k+1)}\,\overline{P}_t\,\hat{C}^{(k+1)\prime}\,\mathrm{Ind}_t^{\mathrm{NA}} + \left(I_n - \mathrm{Ind}_t^{\mathrm{NA}}\right)\,\hat{R}^{(j)}\,\left(I_n - \mathrm{Ind}_t^{\mathrm{NA}}\right) \tag{29}$$

Then, a new step of the algorithm can start from $\left\{\left(\hat{\Lambda}^{(k+1)}, \hat{R}^{j+1}, \hat{A}^{(k+1)}, \hat{Q}^{(k+1)}\right)\right\}$.

A.2 Reconstructing aggregate measures from sample estimates

As anticipated above, the estimation algorithm is applied to H subsamples, $\left\{Y^{(h)}\right\}_{h=1}^{H}$, with constant number of firms (but variable number of flows) of the original dataset[19]. As a result, we obtain H different estimates of the main parameters and latent factors

[16] The index -1 denotes the matrix of lagged factors.

[17] In line with the notation introduced above, $\mathrm{Ind}_t^{\mathrm{NA}}$ denotes a diagonal matrix with ones when the corresponding cross element is available in the cross section t and a zero when it is not.

[18] The matrix $\overline{Z}_{gc}$ by juxtaposition of the restrictions of $\overline{Z}$ to the global and block-specific components.

[19] With a slight abuse of notation we use the same symbols (an index between parenthesis) to indicate partial estimates of the parameters (subsection A.1) and estimates or quantities referred to different sample of the original dataset (A.2).

of equation (1). The family of estimates of the global, $\left\{\hat{f}_t^{(h)}\right\}_{h=1}^{H}$, and destination-specific $\left\{\hat{d}_{c,t}^{(h)}\right\}_{h=1}^{H}$ factors can be used to recover the final estimates by simple average:

$$\hat{f}_t = \frac{1}{H}\sum_{h=1}^{H} \hat{f}_t^{(h)} \quad \text{and} \quad \hat{d}_{c,t} = \frac{1}{H}\sum_{h=1}^{H} \hat{d}_{c,t}^{(h)}$$

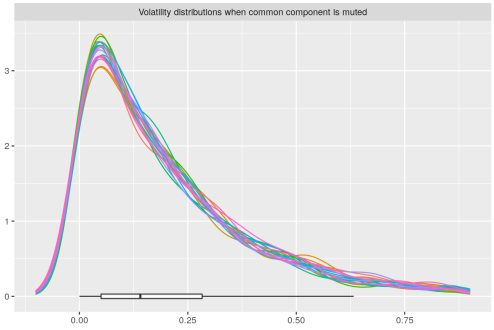
Analogously, we use the outcomes from the H samples to compute the three distinct components of the aggregate volatility. For the global component (σ_g) we have:

$$\sigma_g = \frac{1}{H}\sum_{h=1}^{H} \sigma_g^{(h)} = \frac{1}{H}\sum_{h=1}^{H} \mathrm{Var}_T\left(\sum_j \omega_{j,t} \cdot \hat{\lambda}_j^{(h)} \cdot \hat{f}_t^{(h)}\right) .$$

Confidence intervals for the averages across samples are calculated as follow. Let us denote with $\mathbf{x} = \left(x^1, \ldots, x^N\right)$ the vector of sample estimates of the measure x. We define $\overline{x}$ as the simple average of the elements of $\mathbf{x}$ and approximate the distribution of the deviation from the true means $\delta = \overline{x} - \mu$, by the empirical distribution $\delta^* = \overline{x}^* - \overline{x}$, where $\overline{x}^*$ are the averages of random samples $\mathbf{x}^*$ of elements of $\mathbf{x}$ taken with replacement (equally sized with the same size of $\mathbf{x}$). Once the empirical distribution of δ^* is computed, given a confidence level α, the quantile-thresholds for the values $\frac{1-\alpha}{2}$ and $\frac{1-\alpha}{2}$, defines respectively the lower and upper bound of the confidence interval.

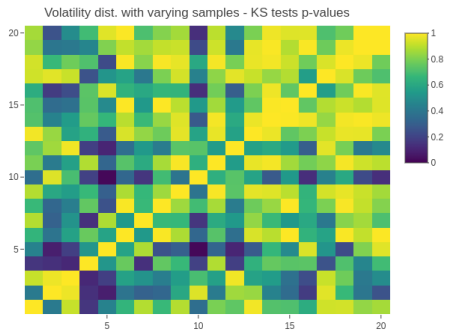
A.3 Firm-level distributions: robustness across samples

Before proceeding with the analysis of the microeconomics of volatility decomposition, let us spend few lines to explain how it is possible to work on firm-level empirical distributions in a context of data subsampling. Instead of taking simple across sample averages and then computing the empirical distributions, we first test the equivalence of distributions of the component (or other related micro-statistics) generated running the estimation method on the different data subsamples. On this topic, figure 14 offers an example: the distributions of volatility with muted global components can be compared both by visual inspection and analysing the heatmap of the statistics of the pairwise Kolmogorov-Smirnov tests. With a good degree of significance, we can treat those empirical distributions as coming from the same population distribution, this being true for all the relevant variables we will consider in the following. This serves our analysis in two ways: first, confirming that the estimation of the micro-level components of the decomposition are coherent and unaffected from biases in firms' selection; second, that we can exploit any of the computed empirical distributions avoiding possible distortions of the following analysis.



(a) Gaussian kernel densities compared.[a]

[a] The non-zero probability assigned to negative values is just an artifact of the fit. A more rigorous estimate would not change the main intuition of the distributions' comparison.



(b) Pairwise KS tests.

Figure 14: Figure (a): a visual comparison of the fitted distribution for the volatility with muted global components. Figure (b): heatmap of the p-values of the pairwise Kolmogorov-Smirnov test for the empirical distribution of the volatility with muted global components.

B Aggregation under fat-tailed distributions

$$
\begin{array}{ccc}
y_{i,t} = \log\left(\frac{x_{i,t}}{x_{i,t-1}}\right) & \xrightarrow{\text{DFM decomposition}} & \mathrm{y}^{C}_{i,t} + \mathrm{y}^{D}_{i,t} + \mathrm{y}^{I}_{i,t} \\
\Vdots\ \text{Same source} & & \vdots\ \text{Aggregation proxies} \\
Y_t = \log\left(\frac{\sum_i x_{i,t}}{\sum_i x_{i,t-1}}\right) & \xrightarrow{\text{Macro decomposition}} & \mathrm{Y}^{C}_{t} + \mathrm{Y}^{D}_{t} + \mathrm{Y}^{I}_{t}
\end{array}
$$

Diagram B.1

Figure 15: An illustrative diagram of the conceptual linkages between microeconomic and macroeconomic decompositions. On the left, the flow level growth rates and their distribution (top) with the aggregate growth rate (bottom), both stemming from the same empirical source: the levels of trade flows (firm-destination pairs). The dynamic factor model serves as baseline to obtain the micro-level decomposition on the top right corner. Then, through selected aggregation strategies (dashed line) the macroeconomic decomposition is recovered, and the two following steps define a linkage between the aggregate growth rates and the macro-component growth rates.

Diagram 15 illustrates schematically the conceptual connections between the relevant statistical entities of the proposed analysis. When analyzing the macroeconomic decomposition implied by the flow-level estimated DFM, the aggregation strategies mapping the flow-level components into the aggregate equivalents need particular attention. Indeed, the possible avenues connecting the elements on the right of the diagram are manifold, and the optimal strategy needs to be weighed looking at the characteristics of the data upon which the analysis relies. We propose here an overview of some recent methods outlined in Bottazzi et al. (2019) — contextualized to our framework — that can complement the results and insights from the more widespread weighted aggregation.

Denoting with $\{x_{i,t}\}$ the level of trade flows we can write the logarithmic growth rate as

$$Y_t = \log\left(\frac{\sum_i x_{i,t}}{\sum x_{i,t-1}}\right) = \log\left(\frac{\sum_i x_{i,t-1}e^{y_{i,t}}}{\sum x_{i,t-1}}\right) \overset{\mathrm{E}}{=} \log\left(\frac{\sum_i x_{i,t-1}\mathrm{E}\left[e^{y_{i,t}}\right]}{\sum x_{i,t-1}}\right) \tag{30}$$

Thus, the problem is reduced to the optimal characterization of the cumulant generating function (CGF) and then the best approximation given the distributional properties of the micro-level growth rates. Noticing that the equation (*) decomposes by construction $y_{i,t}$ in three orthogonal components and that the CGF of the sum of three independent random variables, we can proxy the $\mathrm{E}\left[e^{y_{i,t}}\right]$ as the product of the CGF associated to the three components. Hence, we are induced to consider three different approximation of the CGFs in order of increasing accuracy[20]. For a generic variable Z_t with realizations $\{z_{i,t}\}$:

$$\begin{aligned}
\mathrm{E}\left[e^{Z_t}\right] &\simeq e^{\mu_{Z,t}} \\
\mathrm{E}\left[e^{Z_t}\right] &\simeq e^{\mu_{Z,t}+\sigma^2_{Z,t}} \\
\mathrm{E}\left[e^{Z_t}\right] &\simeq e^{(\overline{x}/x_{i,t})^{2\beta_t}(\overline{C}_{2,t}-C_{2,t})/2} \cdot e^{\mu_{Z,t}+\sigma^2_{Z,t}}
\end{aligned}$$

While the first two formulas are subsequent terms in the series of CGF, the third, relying on more sophisticated theoretical ground, introduces a size dependent correction term accounting for the so called scaling relation between the variance of the growth rates and their size (this subject is thoroughly analyzed in industrial dynamic studies such as Bottazzi and Secchi, 2006). For what concerns our analysis, it is relevant to flag that once the dependence on size is introduced then, first, the three terms of the GR decomposition cannot be disentangled completely because the relative correction terms are absorbed in the sum, and, second, the size-variance relation is ruled by the series of parameters β_t. Hence, when deploying the finer version of the CGF decomposition one has to assess the magnitude of the scaling effect and evaluate the impact of the correction term in the context of application. For the case of interest, table 6 shows that the span of the values of the betas is large and consistently far from the characteristic value of -0.20 that it take when the growth rates are analyzed at the firm rather than at the flow level. With the just exposed caveats, we can test the validity of the examined approximations for the series of the flow level growth rates $y_{i,t}$. In order to understand the appropriateness we compare the approximated aggregation with the actual aggregate growth rate Y_t, in terms of correlation and variance ratio. We notice that, looking at table 7, the Spearman rank correlation suggests that all the three approximations display a significant level of correlation decreasing for increasing accuracy[21], whereas in term of relative variance the the second and the third approximation are the closest to the actual aggregate. Finally, **??** offers a comparison between different aggregation strategies for the three components of the model's main equation.

[20] For the sake of the synthesis, we cannot provide all the details that are thoroughly discussed in Bottazzi et al. (2019)

[21] This is explained by two main consideration, first, A_1 outperform A_2 for a sophisticated mechanism of compensation explained in (Bottazzi et al., 2019) and, second, A_2 outperform A_3 because the impact of beta is probably not relevant or has a contrasting effect.

	Min	1st quartile	Median	3rd quartile	Max
Common	-0.1005	-0.0733	-0.063	-0.0538	-0.02
	(-0.1018,-0.0995)	(-0.0744,-0.0726)	(-0.0635,-0.0626)	(-0.0543,-0.0532)	(-0.0241,-0.0158)
Dest. specific	-0.1	-0.0755	-0.0668	-0.0576	-0.0317
	(-0.1016,-0.098)	(-0.0764,-0.0748)	(-0.0678,-0.0662)	(-0.0581,-0.0571)	(-0.0336,-0.0299)
Idiosyncratic	-0.0564	-0.0455	-0.0411	-0.0358	-0.0211
	(-0.0572,-0.0558)	(-0.0459,-0.0451)	(-0.0415,-0.0408)	(-0.0363,-0.0354)	(-0.0218,-0.0204)

Table 6: Quantiles of the distribution along the time dimension of βs for the three components

	A_1	A_2	A_3
Corr.	0.8564	0.7414	0.6981
Rel. Var.	0.0064	1.11	0.98

Table 7: Correlations and relative variance of the three aggregation proxies with respect to the log growth rate of the aggregate.

Volatility Share

Dynamic Weighted Sum			Constant Weighted Sum		
Global	Dest. Spec.	Idio.	Global	Dest. Spec.	Idio.
0.254	0.218	0.724	0.328	0.230	0.734
(0.207,0.302)	(0.184,0.258)	(0.691,0.749)	(0.259,0.372)	(0.208,0.250)	(0.688,0.760)

Approximate CGF			Corrected CGF (Bottazzi et al. 2020)		
Global	Dest. Spec.	Idio.	Global	Dest. Spec.	Idio.
0.471	0.272	0.775	0.297	0.215	0.895
(0.446,0.502)	(0.251,0.305)	(0.745,0.796)	(0.281,0.319)	(0.198,0.233)	(0.838,0.924)

Table 8: Aggregate volatility shares for the global, destination-specific and idiosyncratic components. Aggregation based on weighted sums (top panels), with constant or dynamic flow-specific loadings (along with respectively Di Giovanni et al., 2018; Kramarz et al., 2020), are compared with aggregations accounting for fat-tails and variance-size scaling relations.