OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 0305-9049
doi: 10.1111/obes.12639

Factoring in the Micro: A Transaction-Level Dynamic Factor Approach to the Decomposition of Export Volatility

MATTEO BARIGOZZI,† ANGELO CUZZOLA,‡ MARCO GRAZZI§ and DANIELE MOSCHELLA‡

†*Department of Economics, Università di Bologna, Bologna, Italy*
‡*Institute of Economics and L'EMbeDS, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà 33 Pisa, Italy (e-mail: angelo.cuzzola@santannapisa.it)*
§*Department of Economic Policy, Università Cattolica del Sacro Cuore, Milan, Italy*

Abstract

This paper analyzes the export volatility sources estimating a dynamic factor model on transaction-level data. Utilizing an exhaustive dataset of French export transactions from 1993 to 2017, we reconstruct the latent factors space associated with global and destination-specific macroeconomic shocks through a Quasi-Maximum likelihood approach which allows accommodating both the high share of missing values and the high dimensionality of the microeconomic time series. The estimated parameters are then used to derive a volatility decomposition of the aggregate and firm-level export growth rates, highlighting structural spatial patterns and the role of geographical diversification in mitigating export risks.

I. Introduction

Finely disaggregated data have fostered a fast-growing body of research on microeconomic trade flows and their influence on firm-level and aggregate growth rates, with many empirical studies exploiting the granular information to investigate possible avenues of the micro-to-macro channel. However, as pointed by Armenter and Koren (2014), a proper model of the trade activities of firms and countries should always take into account some

JEL Classification numbers: C38, L25, F14.

This paper benefited from helpful comments at several conferences, including IAAE 2021 annual conference and CAED 2021. We are also indebted to Julien Martin and Marco Cococcioni for their insightful comments, and grateful to two anonymous referees for their valuable suggestions and comments. This work has been partly supported by the European Commission under the H2020, GROWINPRO, Grant Agreement 822781 and by the Italian Ministry of Education and Research under the PRIN-2017 Programme, project code 201799ZJSN, PRIN2022 (GREESCCO), project code P2022LP43N and PRIN2022, project code 2022ACJ2WS. 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 disclaimers apply.

[Correction added on 28 September 2024, after first online publication: The author order list has been changed in this version.]

key structural properties of transaction-level datasets: the pervasive sparsity, intended as the number of zeros in trade flows at the microeconomic level, and the skewness of cross-sectional distributions. Those aspects have fundamental implications for the design and interpretation of trade models and their stylized facts, gaining even more relevance when the focus lies on the dynamics of economic interactions. This is certainly the case of the volatility models based on the micro-level decomposition of the growth rates (Kelly *et al.*, 2013; di Giovanni *et al.*, 2014; Kramarz *et al.*, 2020).

In this context, our paper presents a novel approach to the decomposition of the growth rates, based on the identification and estimation of global and local co-movements for high-dimensional time series with arbitrary patterns of missing data. The method requires the estimation of an approximate dynamic factor model (DFM) with a block structure (BDFMs in short, see e.g. Hallin and Liška, 2011; Moench *et al.*, 2013) for the growth rates of export value at the firm-destination level for the universe of French exporting firms. DFMs endowed with a block structure feature both global and local factors to capture, respectively, the co-movements common to all the time series and co-movements within blocks of series. Thus, associating a block to each served destination, we provide an additive decomposition of the growth rates into three orthogonal terms: the global component, embodying the contribution of a global latent factor and the related loading, the destination-specific component, endowed again with a related loading and capturing the influence of the local factor associated to the target destination, and an idiosyncratic irreducible component.

We estimate the model via a Quasi-maximum Likelihood (QML) approach implemented through the Expectation Maximization (EM) algorithm (Watson and Engle, 1983; Doz *et al.*, 2012), in line with the applications proposed by Bańbura and Modugno (2014). We extend their work giving the explicit derivation of the M-step for models with a block structure and proposing a suitable initialization procedure based on the sequential least square estimator of Breitung and Eickmeier (2015), recovering consistent estimates of the global and local factors, the related loadings and the idiosyncratic terms for highly dimensional time series in presence of missing values. Macroeconomic applications of a similar estimation technique based on a block structure of the data are Coroneo *et al.* (2016) and Delle Chiaie *et al.* (2022) (see below for details).

The approach described above allows to improve the existing decompositions of export volatility along three main dimensions. First, in contrast to di Giovanni *et al.* (2014) and Kramarz *et al.* (2020), the macroeconomic determinants of volatility – driven by global or destination-specific factors – are explicitly modelled as autoregressive stochastic processes, making efficient use of the information available at the microeconomic level. Second, DFMs capture the relevance of macroeconomic shocks, not only *per se*, but also as drivers of heterogeneous responses at the firm level. Indeed, the global and destination-specific components are defined as the product between the latent factors and the so-called factor loadings, that is, numerical time-invariant coefficients specific to each firm-destination cell. These parameters represent the elasticity of firm-destination growth rates to common movements encoded in the global and local factors. Third, this exercise can be carried out by increasing the time frequency of the data without restricting the dataset to persistent exporters only, as the estimation technique efficiently tames the increasing number of missing values. Here, we use quarterly data on firm-destination transactions for around 144,000 firms exporting to 67 destinations from 1993 to 2017, a period that includes relevant macroeconomic events (e.g. the trade collapse) and different phases of French export cycles. These yearly quarter-to-quarter growth rates allow us to work on long time series, while mitigating the bias of partial-year effects on first-year export growth rates (Bernard *et al.*, 2017).

We contribute to several streams of literature. First, we present a novel application of dynamic factor models to firm-level and transaction-level data. To the best of our knowledge, our work is the first to estimate a DFM on such disaggregated data, which allows us to identify the interactions between macroeconomic factors and heterogeneous agents. Our approach is aligned with the literature on block DFMs, postulating global and local factors to capture hierarchical correlation structures within economic and financial datasets. Existing papers in this area have applied similar models to assess the relative importance of world, regional, and country-specific factors on countries' business cycles (Kose *et al.*, 2012; Mumtaz and Surico, 2012; Miranda-Agrippino and Rey, 2020), to decompose the variance of commodity price indexes taming the local cross-correlation within groups (Delle Chiaie *et al.*, 2022), and to impose restrictions on the loadings of nominal and real variables (Coroneo *et al.*, 2016). However, our application differs in that it requires the estimation of a cross-section that is four orders of magnitude larger and accommodates a high share of missing values, reaching approximately 70% of the dataset. Thanks to the methodology we propose, the incomplete firm-destination time series need not be excluded or imputed and the growth rates' comovements can be estimated by exploiting all available information overcoming the concerns on the varying distribution of missing values at different time steps.

Second, acknowledging the prominent role of trade flows in contributing to GDP volatility (see di Giovanni and Levchenko, 2009 and the Figures 1a,b), we provide new quantitative estimates of the *granular* component of the volatility of French exports, thanks to the decomposition of transaction-level growth rates. Starting from the seminal work by Gabaix (2011), a rich stream of literature has shown that in a *granular economy*, with a fat-tailed firm size distribution, idiosyncratic shocks to individual firms explain a significant part of the aggregate movements (Acemoglu *et al.*, 2012; Carvalho and Gabaix, 2013; Carvalho and Grassi, 2019).[1] Those effects become increasingly relevant in international trade (di Giovanni and Levchenko, 2009; di Giovanni *et al.*, 2014; Di Giovanni *et al.*, 2018; Bricongne *et al.*, 2022), whereby the exporters' size distribution is even more skewed (among the others, Bernard *et al.*, 2009, 2016). This literature typically represents the growth rate of exporters as the sum of orthogonal terms, including macro shocks, estimated statically from each cross-section, and micro perturbations. We will refer later to these as static orthogonal decomposition models (SODMs). In such a context, we use the estimated BDFM to provide a new decomposition of the aggregate volatility into global, destination-specific and idiosyncratic components. Our results provide a novel perspective on the role of granular vs. common effects in aggregate fluctuations. We show that common shocks to firm-destination export growth rates

[1] First highlighted for firms' size and aggregate GDP, that intuition applies to several economic phenomena (see e.g. Mésonnier and Stevanovic, 2017; Amiti and Weinstein, 2018, for the macroeconomic effects of microeconomic shocks in the banking sector).

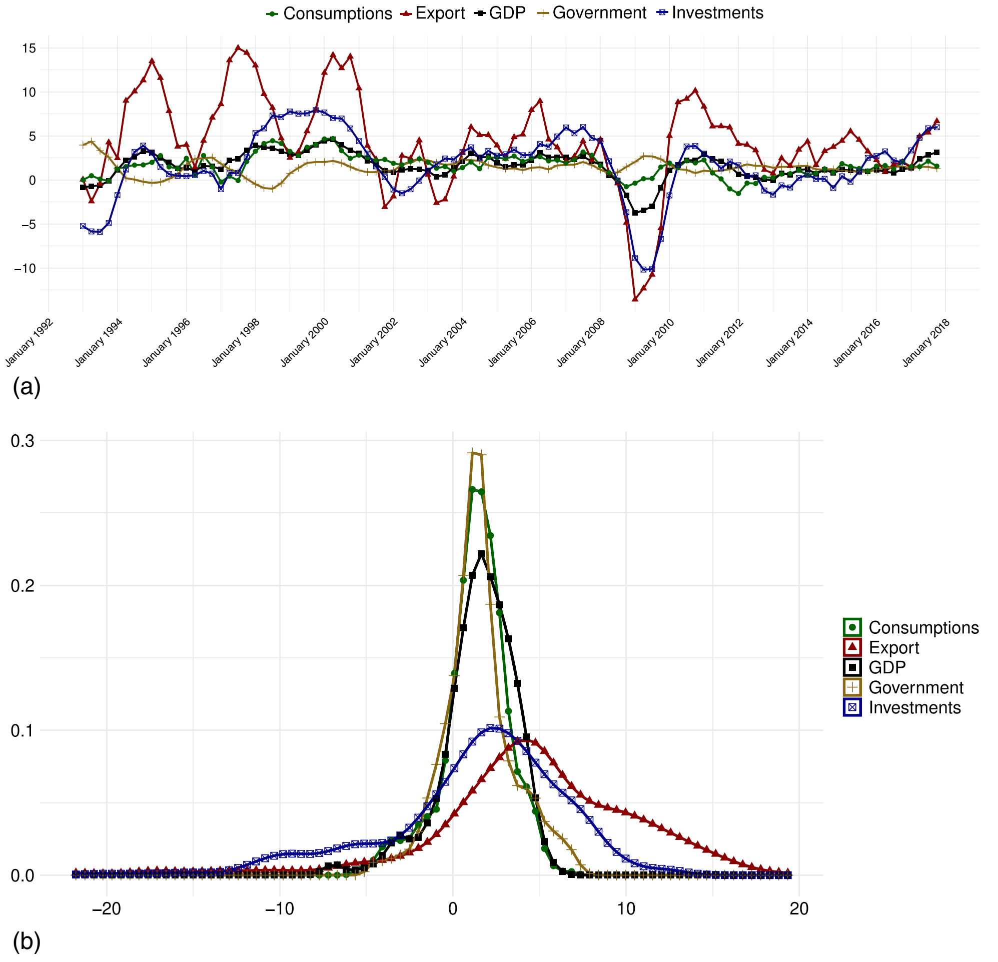


FIGURE 1. The growth rates of the GDP and its components. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: Our elaboration on the OECD Main Economic Indicators (MEI) dataset, including quarterly deseasonalized growth rates (measure GYSA) for the GDP and its main components: GDP [NAEXKP01], Export [NAEXKP06], Investments [NAEXKP04], Governments [NAEXKP03], Consumptions [NAEXKP02]. (a): time series of France's quarterly deseasonalized growth rates. (b): Gaussian kernel estimates of the pooled distribution of quarterly GRs for France, Italy, Germany and Spain from 1993 to 2018. (a) Time-series. (b) Kernel density estimations.

are significantly more volatile than common components identified via SODMs: they can generate significant variations in the micro growth rates because firms' reactions are highly heterogeneous. This mechanism has important macroeconomic implications. When decomposing aggregate volatility, the significant variation in the firms' reactions induces relevant changes in the estimates of the common component during different phases of the business cycle. In particular, the common component of the aggregate volatility surged during the Great Trade Collapse, matching in magnitude the idiosyncratic one. In normal times, our firm-destination specific component accounts for the most significant part of aggregate volatility, similar to standard models. This different representation of the time-varying behaviour of the volatility is due to the adopted methodology based on the Kalman smoother, which aggregates data in an optimal way (in mean squared sense) by taking cross-sectional and dynamic weighted averages of all observed time series. This is fundamentally different from taking sectoral (or destination-specific) averages to identify macroeconomic effects and isolate the granular component (Gabaix (2011); Carvalho and Gabaix (2013); di Giovanni *et al.* (2014); Bricongne *et al.* (2022)) since by dynamically aggregating the data we allow for leading-lagging relations among time series which account for feedback effects from the level of global and local factors to the data and vice versa (see Figure).

Third, using the same decomposition, we analyze the volatility at the firm level providing new insights on the volatility-diversification nexus. Several contributions find a dampening effect of diversification on volatility (see e.g. Herskovic *et al.*, 2020). In particular, firms selling multiple products to multiple destinations are those responsible for the largest share of a country's export flows (Eaton *et al.*, 2004; Bernard *et al.*, 2012) and they could reduce their volatility by diversifying their portfolio (di Giovanni *et al.*, 2014; Kramarz *et al.*, 2020). We measure the effects of the three distinct components on the firms' volatility distribution, showing how the global and the destination-specific terms generate a significant part of the risks inherent to export growth, even though the impact of the idiosyncratic non-reducible components is relatively higher. The decomposition is then used to understand how and to what extent geographical diversification strategies help dampen export volatility overall and the single components on their own. We find that while firms that diversify across destinations succeed in mitigating the risks associated with the macroeconomic cycle, an inverted U-shaped relationship between diversification and idiosyncratic volatility suggests that the same strategies do not mitigate idiosyncratic risks until a certain level of diversification is reached.

The remainder of the paper is structured as follows. Section II introduces the model and gives a brief and concise description of the methodology. Section III offers a bird's eye view of the dataset characteristics, focusing on sparsity, the distribution across destinations, and some firm-level statistics. Section IV presents the reconstruction of the latent factor space and the volatility decomposition at the aggregate and firm levels. Section V concludes.

II. Model and methodology

This work provides an econometric framework to identify different sources of shocks affecting international trade flows. The methodology guarantees a high degree of flexibility as the structure that we impose *a priori* is kept to a minimum and is ultimately derived from the fundamental composition of the disaggregated transaction data (see Figure 3). The approach allows for identifying global and destination-specific components of the growth rates of the exports and their influence on aggregate and firm-level volatility. We build upon a widespread class of models that typically represent the growth rate of exporters as the sum of orthogonal terms, including macro shocks, estimated statically from each cross-section, and micro perturbations (see e.g. di Giovanni *et al.*, 2014; Kramarz *et al.*, 2020; Bricongne *et al.*, 2022). In line with these SODMs, we recover the growth rates of the export sales directed to a given destination as the sum of three parts: two terms accounting for the macroeconomic effects of a global and a destination-specific component and a third microeconomic component that is specific to firm-destination pairs.

The proposed methodology improves upon existing SODMs in the characterization of the macroeconomic effects: we assume that firms-to-destination trade varies in response to macroeconomic movements common to all the firms (global) or affecting only the firms exporting to a given destination (destination-specific). Those movements are represented by latent global and destination-specific factors and come endowed with their own dynamic specification. Moreover, the model is considerably richer since the dynamics of the microeconomic growth rates are driven by heterogeneous responses to the factors through elasticities (factor loadings) that depend on the specific match between firms and destinations. In practical terms, this decomposition is achieved by estimating a dynamic factor model with a block structure or block-DFM, induced by geographical export patterns. Once estimated, the growth rate decomposition is scaled up at different levels of aggregation, recovering the volatility decomposition for the total export, the export to each specific destination, and the firms' export.

Our proposed application of DFM also improves upon existing techniques in yet another way: the estimation of factors based on the Kalman smoother is fundamentally different from the estimation of fixed-effect at each cross section as in Gabaix (2011), Carvalho and Gabaix (2013), di Giovanni *et al.* (2014), Bricongne *et al.* (2022) since by dynamically aggregating the data we allow for leading-lagging relations among time series which account for feedback effects from the level of global and local factors to the data and vice versa. Consider Figure 2, representing the information flow and the implied dynamics within the two frameworks. For DFMs, along the vertical dimension, information flows are bidirectional. In a two-step sequential estimation procedure, the estimated factors are used to determine the loadings that are then exploited to update the factor estimates until convergence. This interplay between the horizontal and vertical dimensions allows for a full-fledged dynamic decomposition and a much more efficient handling of the available information.

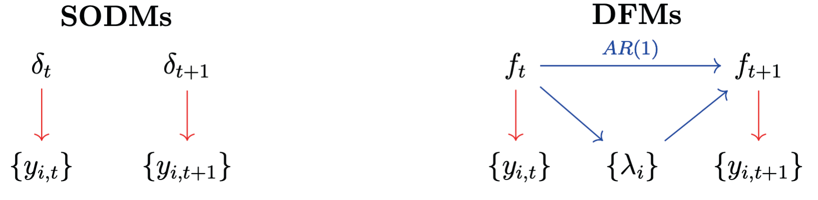


FIGURE 2. Information flows for SODMs and DFMs. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: The diagrams compare the information flowing directly (red arrows) or indirectly (blue arrows) in orthogonal decomposition models (SODMs) and dynamic factor models (DFMs) for a generic time series of growth rates $\{y_{i,t}\}$ (with i running over the number of flows). The symbols f_t and λ_i denote, respectively, a common latent factor, whose dynamics are modelled as an $AR(1)$ process, and the associated loadings. Common components in SODMs are estimated using the information from single cross-sections only, while in DFMs sequential time steps jointly concur to the formation of factors and loadings estimates.

Model's equations and estimation

Working with flows of exports at the firm-country level, we denote with $y_{de,t}$ the growth rate of the flow towards destination d by exporter e. We specify a simple model, featuring both macroeconomic shocks and destination-exporter specific shocks, similar to di Giovanni *et al.* (2014). The model is described by the set of equations:

$$y_{de,t} = \lambda_{de} f_t + \rho_{de}\, g_{d,t} + \xi_{de,t}. \tag{1}$$

$$f_t = a_f\, f_{t-1} + u_{f,t}. \tag{2}$$

$$g_{d,t} = a_d\, g_{d,t-1} + u_{d,t}. \tag{3}$$

With equation (1) we postulate the influence on $y_{de,t}$ of one latent factor common to all the flows, f_t, and one $g_{d,t}$ specific to the destination d. Factors are assumed to be orthogonal, and we model the dynamics of each factor as autoregressive processes of order one.[2] The loadings λ_{de} and ρ_{de} reflect the response to the factor-related shocks, $u_{f,t}$ and $u_{d,t}$. Such loadings are specific to each firm-destination pair, thus embodying the possible heterogeneous response of exporters to global and destination-specific shocks. Notice that, while the complete model features one global factor and D destination factors, if we naively restrict the model to the space spanned by the series directed to a given destination the model becomes a simple DFM with two factors, where the variance explained by the second factor (the destination-specific one in the complete representation) is residual with respect to that explained by the first one.

Dynamic factor models have an established tradition in macroeconometrics. Their early applications were based on the exact factor structure assumption that all cross-correlations in the data could depend on a few common factors while the idiosyncratic noise remained cross-sectionally uncorrelated (Sargent and Sims, 1977; Quah and Sargent, 1993). Unfortunately, this becomes an unrealistic hypothesis in the case of interest to our research question, where the dataset's cross-sections are very large, favouring the emergence of idiosyncratic cross-sectional correlations. Estimation of such approximate factor structures has been originally proposed in Stock and Watson (2002) and Bai (2003) using classical principal component analysis and only global factors. Estimation of local factors via principal components has been studied by Onatski (2012), Breitung and Eickmeier (2015), and Freyaldenhoven (2022), among others. These works show how in presence of a factor structure, a high-dimensional dataset with high-dimensional blocks is precisely what allows to consistently estimate the global and local factors and their loadings even in presence of (limited) cross-correlation among idiosyncratic components, thus transforming a potential curse of dimensionality into a blessing.

More recently, Doz *et al.* (2012) proposed to estimate DFMs via QML implemented via the EM algorithm, a method originally proposed by Watson and Engle (1983), Shumway

[2]The estimation of a model with higher order AR processes and with multiple common factors or multiple destination-specific factors is possible and technically feasible. Appendix C presents the volatility decomposition obtained with AR processes of order four, but we keep fixed the number of factors to avoid excessive parameter proliferation.

and Stoffer (1982), and Quah and Sargent (1993) for small-size state space models. This approach is particularly suited to deal with missing data (Mariano and Murasawa, 2003; Marcellino and Schumacher, 2010; Bańbura and Modugno, 2014), yet it is widely used generally on financial and economic applications (see e.g. Coroneo *et al.*, 2016; Delle Chiaie *et al.*, 2022). An alternative factor-based approach leveraging the block structure of the data is the hierarchical factor model considered in Moench and Ng (2011) and Moench *et al.* (2013). Notwithstanding the differences in its general formulation, the decomposition approach suggested by those authors aligns with ours for the application proposed in this paper. However, their model may present challenges in aligning with the volatility decomposition literature and, as far as we are aware, it is not endowed with estimation procedures well-suited for datasets with large cross-sections and missing values.[3,4]

Hence, building upon the literature on EM estimation of DFMs, we extend the framework dealing with missing values and block structures simultaneously. We also propose an *ad hoc* initialization procedure, based on a least square sequential estimator proposed by Breitung and Eickmeier (2014, 2015) (see the Appendix A for the details). In addition, notice that, since we work with huge cross-sections (approaching an order of magnitude of $\sim 10^6$ units) and tens of factors, the estimation entails a consistent memory overload, which cannot be tamed without an efficient handling of the sparse matrices of input data and loadings.

The EM algorithm consists of an iterated estimation procedure that converges towards the QML estimates of the parameters in a sequence of steps. We briefly recall here its fundamentals, referring to Appendix A for a detailed description. Adopting the synthetic notation, labelling the factors as F, the other parameters[5] as θ and the matrix of data as Y, we write the joint log-likelihood of data and factors as $l(Y, F, \theta)$, and iterate the following two steps:

1. Given an estimate of the parameters $\theta^{(k-1)}$, we derive the factors $F^{(k-1)}$ (jointly with their covariance) via the Kalman smoother, then compute the expected joint log-likelihood of Y and F:

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

   where Ω_T denotes the σ-field generated by Y and the expectation is taken using the conditional distribution of the factors given Ω_T and the estimated parameters $\theta^{(k-1)}$.
2. Obtain an update of the parameters by solving:

$$\theta^{(k)} = \arg\max_{\theta} L\left(\theta; \theta^{(k-1)}\right).$$

[3]In fact, even if Moench *et al.* (2013) is in general specified in a dynamic way, in order to identify the factors, the applications proposed in those papers are restricted to the case of static loadings. Therefore, as pointed out by Moench and Ng (2011), the responses of variables to shocks to local factors in a given block can differ only to the extent that their exposure to the block-level factors differs, in complete agreement with the proposed model.

[4]Concerning comparability, rewriting with our notation equation (5) from the section 'Related Work' Moench *et al.* (2013), the model main equation would become $y_{de,t} = \lambda_{de}^{\mathrm{MNP}} \lambda_{d}^{\mathrm{MNP}} f_t + \lambda_{de}^{\mathrm{MNP}} u_{d,t}^{\mathrm{MNP}} + \xi_{de,t}^{\mathrm{MNP}}$ (where the suffix MNP is used when the parameters are not directly comparable to ours). Such an equation cannot be derived from a model *à la* di Giovanni *et al.* (2014), as for example the destination factor is not influencing the growth rates directly via its fluctuations but through its innovation $u_{d,t}^{MNP}$.

[5]Namely, the loadings matrix, the VAR coefficients and the relevant variance−covariance matrices.

This cycle defines a sequence of increasing log-likelihood values

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

and stops when an appropriate convergence condition is fulfilled. While typically the algorithm is initialized by principal component estimates of loadings and factors, in order to account for block-specific factors, we propose to initialize it using the following iterative procedure: first, missing values are imputed using time-series medians and moving average smoother, then a sequential least square estimator proposed by Breitung and Eickmeier (2014, 2015) is applied on the 'completed' matrix to obtain the block-by-block parameters initialization. In this respect, whereas applied on imputed data, we can exploit the well established asymptotic properties of the initializing estimator (see propositions 2 and 3 of Choi *et al.*, 2018).

Throughout, we employ a misspecified likelihood where the idiosyncratic components are treated as if they were both cross-sectionally and serially uncorrelated and normally distributed. This makes estimation fast and easy, and allows to have analytical expressions for the solutions at the maximization step. Nevertheless, it can be shown that, as both the total number of series, $\sum_{d=1}^{D} n_d$, and the sample size, T, grow to infinity, the consistency and efficiency of the estimated loadings and factors are not affected by such misspecifications. Moreover, the estimated factors are likely to be more efficient than those recovered by principal component analysis. We refer to Bai and Li (2016) and Barigozzi and Luciani (2019) for more details on the asymptotic properties of the estimators. Furthermore, we stress the robustness of the methodology to deviations from Gaussianity of the idiosyncratic terms and the factors' innovations, collecting Monte Carlo evidence on the finite sample properties under this and other misspecifications (see Appendix B) and joining several empirical and numerical applications to leptokurtic or asymmetric distributed data (see e.g. Reis and Watson, 2010). These results justify the application to the growth rates of the export transactions whose distributional properties will be explored in the following section. Finally, notice that the estimation is performed on standardized and demeaned series, then the estimated values are remapped to the original scales before proceeding with aggregation.

Volatility estimates and decompositions

Our approach aims to the identification of idiosyncratic and macroeconomic shocks to export sales growth rates and to the estimation of their impact on the volatility of the aggregate. While the model outlined in (1)–(3) differs from the existing identifying methodologies adopted by Gabaix (2011), di Giovanni *et al.* (2014) and Kramarz *et al.* (2020), we draw from their aggregation strategy for the mapping between the microeconomic decomposition and the macro outcomes.[6] We report here a quick overview of the main lines.

[6] An alternative weighting scheme along the lines of Bricongne *et al.* (2022) is explored in Appendix E. For a thorough review of the possible strategies to recover the aggregates from microeconomic flows decompositions, see Amiti and Weinstein (2018).

Once estimated, the model main equation provides a decomposition of the logarithmic growth rates of each exporter-destination cell

$$y_{de,t} = \widehat{\lambda}_{de}\,\widehat{f}_t + \widehat{\rho}_{de}\,\widehat{g}_{d,t} + \widehat{\xi}_{de,t}. \tag{4}$$

These estimates will be used to assess the impact of any of the terms (or a combination of two) on the aggregate fluctuations. In order to recover the aggregate we will make use of size-based weights, encoding the share of the single exporter-destination cell within a given aggregation level:

$$\omega^{\text{agg}}_{de,t} = \frac{y_{de,t-1}}{\sum_{d,e} y_{de,t-1}}. \tag{5}$$

While summing up the contribution of the single flows to the aggregate, these weights can be chosen to be fixed (di Giovanni *et al.*, 2014) or time-varying (Kramarz *et al.*, 2020). Then one can recover the aggregate time series as:

$$\gamma^{\text{agg}}_{t|\tau} = \sum_{d,e} \omega^{\text{agg}}_{de,\tau} \left(\widehat{\lambda}_{de} \cdot \widehat{f}_t + \sum_d \widehat{\rho}_{de,i} \cdot \widehat{g}_{d,t} + \widehat{\xi}_{de,t} \right), \tag{6}$$

$$\gamma^{\text{agg}}_{t} = \sum_{d,e} \omega^{\text{agg}}_{de,t} \left(\widehat{\lambda}_{de} \cdot \widehat{f}_t + \sum_d \widehat{\rho}_{de,i} \cdot \widehat{g}_{d,t} + \widehat{\xi}_{de,t} \right), \tag{7}$$

where in the first equation we set up the weights to be fixed at a given time step τ, while in the second they are allowed to vary along the time series together with the growth rates components. To construct a proxy of the aggregate volatility, we will work with the variances and standard deviations of the quantities in (6) and (7). In particular, we define the actual or aggregate variance as $\sigma^2_{\text{agg},\tau} = \text{Var}\left(\gamma^{\text{agg}}_{t|\tau}\right)$ and those of the components as:

$$\sigma^2_{\text{glob},\tau} = \text{Var}\left(\sum_{d,e} \omega^{\text{agg}}_{de,\tau} \cdot \widehat{\lambda}_{de} \cdot \widehat{f}_t \right) \qquad \sigma^2_{\text{glob}} = \text{Var}\left(\sum_{d,e} \omega^{\text{agg}}_{de,t} \cdot \widehat{\lambda}_{de} \cdot \widehat{f}_t \right)$$

$$\sigma^2_{\text{dest},\tau} = \text{Var}\left(\sum_{d,e} \omega^{\text{agg}}_{de,\tau} \cdot \sum_d \widehat{\rho}_{de,i} \cdot \widehat{g}_{d,t} \right) \qquad \sigma^2_{\text{dest}} = \text{Var}\left(\sum_{d,e} \omega^{\text{agg}}_{de,t} \cdot \sum_d \widehat{\rho}_{de,i} \cdot \widehat{g}_{d,t} \right)$$

$$\sigma^2_{\text{idio},\tau} = \text{Var}\left(\sum_{d,e} \omega^{\text{agg}}_{de,\tau} \cdot \widehat{\xi}_{de,t} \right) \qquad \sigma^2_{\text{idio}} = \text{Var}\left(\sum_{d,e} \omega^{\text{agg}}_{de,t} \cdot \widehat{\xi}_{de,t} \right).$$

Let us emphasize a few points on the characteristics of these aggregate variances. First, note that the aggregation with fixed weights provides T different estimates of the volatility, depending on the weight selected, so we will consider the time average as the point estimate for the whole time span. Second, the variance of the aggregate cannot be recovered as the simple sum of variances of the components, because of the covariances of the paired terms.

In this respect, non-null covariances between the estimated components might emerge even though orthogonal decomposition models (dynamic or static) assume their independence in population. Following the convention in the literature, throughout the paper we measure volatility as the standard deviation and define the *relative standard deviations* or the *relative volatility* of the global, destination-specific and idiosyncratic components as, respectively, the ratios of the form $\sigma_{\text{glob},\tau}/\sigma_{\text{agg},\tau}$, $\sigma_{\text{dest},\tau}/\sigma_{\text{agg},\tau}$ and $\sigma_{\text{idio},\tau}/\sigma_{\text{agg},\tau}$.

In analogy with the formulas for the aggregation over all the export transactions, it is possible to generalize to different levels of aggregation. In this paper, we work both with destination-specific and firm aggregates. The former is obtained by aggregating the series targeting a specific destination d, thus the weighting can be restricted to the set, I_d, of flows targeting d,

$$\omega_{e,t}^{(d)} = \frac{y_{de,t-1}}{\sum_{e \in I_d} y_{de,t-1}}. \tag{8}$$

On the same line, firm-specific volatility can be obtained as the SD associated to the sum of the transactions of each exporter directed to any destination. Formally, taking the portfolio of the destinations for the exporter e (I_e), the weights become

$$\omega_{d,t}^{(e)} = \frac{y_{de,t-1}}{\sum_{d \in I_e} y_{de,t-1}}. \tag{9}$$

Both for destination and firm-level aggregates the same observations on the dynamic and static weighting apply.

Volatility estimates uncertainty and dataset subsampling

The procedure for the computations of the aggregate volatilities composes of two main steps: the estimation of the elements of the growth rate decomposition and the aggregation. This raises the issue of understanding how the various sources of uncertainty compose to the definition of the confidence intervals. di Giovanni *et al.* (2014), in their Appendix C, derive confidence intervals under the premise that the primary elements of the decomposition, namely the common and idiosyncratic terms of the growth rates, are observed (and thus known) rather than estimated. They then account for the aggregation uncertainty by constructing confidence intervals using both theoretical asymptotic distributions and bootstrapped SEs.

In line with this, aiming to underscore the differences of the growth rate decompositions only, we presented our results in Figure 8 using theoretical confidence bands as derived by di Giovanni *et al.* (2014) (see Appendix C therein).

Nonetheless, we acknowledge that disregarding the uncertainty in the decomposition's estimation can be a limitation. To address this, in our work, we present also estimates of the aggregate volatility which are composed from estimates obtained from dataset's subsamples.

In brief, we estimate the model on random subsamples of the original dataset. We first build a predefined number of reduced datasets (H), selecting at random a fixed number of firms (N_h) and keeping all the time-series associated with those firms. Then we apply the estimation procedure to each reduced dataset, obtaining H different estimates of the factors and the aggregate volatility coming from the decomposition (1). The final estimates for the factors and the aggregate volatility are thus averages of the H factors estimates and H volatility estimates and come along with the relative confidence intervals.[7]

III. Data and stylized facts

To estimate the model outlined above 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 a monthly basis for each year from 1993 to 2017 for all French exporters. A unique official identification number identifies each exporter (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). Our analysis relies on export values at the firm-country level. We start by applying standard cleaning methodologies described in Bergounhon *et al.* (2018). They include 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 1,000 euros (or 1,000 kg) was not compulsory before 2010, we opted for the deletion of all the transactions below the threshold before and after 2010. In total, we dropped around 1.5 millions of firm-product-destination-month tetrads per year, accounting for a tiny fraction of 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, which constitute the units of our analysis. We then aggregate monthly data into quarterly data and transform the panel of transactions into a matrix of time series, one for each firm-destination pair. Notice that the dataset includes the universe of (legal) transactions and it is originally provided in the so-called 'long format', where the id of each firm gets repeated as many times as the number of transactions in a given year. When creating a panel, we artificially generate missing values which we proceed to fill with zeros. 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. Their incidence and distribution need to be analysed in order to proceed with estimation.

The first relevant issue arising in the estimation of the model (1) regards the sparsity of the dataset: 89.99% of all observations are missing. Moreover, available points are unevenly distributed across firm-destination pairs: Figure 4 shows that on a log-log scale the distribution of available information follows a Pareto-like distribution. Within our

[7]Notice that while defining the subsampling parameters a trade-off is at stake. For the weighted sum to be interpreted as an economic aggregate one should select a number of firms not too far from the universe of French firms. On the other hand, the samples should be small enough to exclude excessive similarity which would undermine the sense of the exercise. We opted for a sample dimension of around 70% of the complete dataset (100k firms, when the firms in the dataset are 144k). Accepting these caveats, the results and bands from Figures 6 and 9, Table 2 can be considered informative both on the point estimates and on the combined uncertainty.

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

FirmID	Dest.	Value	Year	Quarter
00215	DE	6,126	1993	04
00215	DE	8,114	1994	02
00215	DE	1,134	1994	03
00215	IT	474,469	1993	03
00215	IT	127,050	1993	04
00215	IT	55,780	1994	03
00215	IT	357,415	1994	04

Time flow

FirmID	Dest.	...	1993.03	1993.04	1994.01	1994.02	1994.03	1994.04	...
00215	DE	...	—	6,126	—	8,114	1,134	—	...
00215	IT	...	474,469	127,050	—	—	55,780	357,415	...

FIGURE 3. Dataset stylized structure. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: 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.

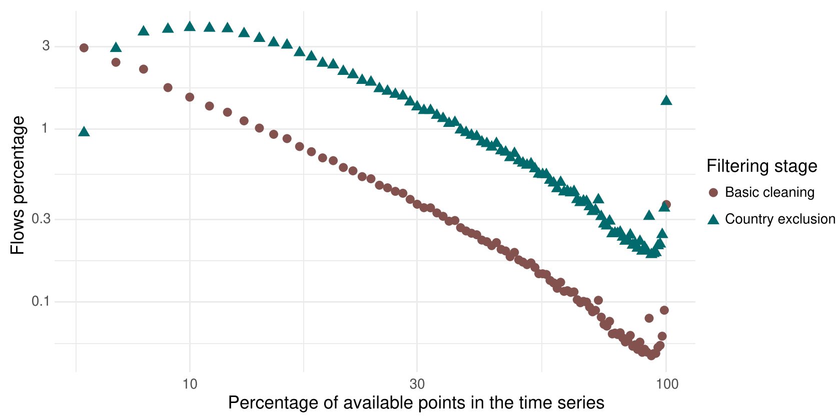


FIGURE 4. Dataset's available information. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: Percentage of active flows (y axis in $\log_{10}$ scale) vs. the percentage of available points in the time span, that is, number of active quarters (x axis in $\log_{10}$ scale). Two curves are proposed to compare the missing value distribution before and after the filtering procedure implemented along the country dimension.

interval span the share of missing values in the cross-section has minor variations over time, with a slightly improving situation in the more recent years. These changes along the time dimension do not pose serious issues for the estimation because the maximum observed spread is around 3%.

The second problem concerns the skewness along the country dimension. Among the 259 countries included in the original dataset, only a few of them are relevant for our analysis. For example, during the considered period only 67 countries report at least 1000 active flows at any cross-section (out of potential 3.2 million of flows).

Time frequency and country restrictions

Given the properties of the dataset, the choice of the optimal time frequency and the selection of destination countries are of strategic relevance. Concerning the frequency, we construct yearly growth rates by taking quarter-to-quarter logarithmic growth rates on quarterly data in an attempt to enhance the standard volatility analysis in two different aspects. First, quarterly data allow us to deal with sufficiently long time series (96 points in yearly quarter-to-quarter growth rates) providing the proper dimensionality for the identification of common and destination factors. Notice that yearly quarter-to-quarter growth rates are also functional in removing at best the seasonality in each time series without adopting additional filtering procedures. 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), which 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. The advantages of higher time frequency for studying the aggregate volatility and its determinants are clear, as evidenced by the most recent studies in the field (see e.g. Bricongne *et al.*, 2022). However, the use of quarterly and monthly data introduces the challenge of intra-year seasonality. By converting the data into quarter-to-quarter logarithmic growth rates, we can partly alleviate the seasonality effect, with some residual seasonality potentially emerging as fourth-order autocorrelation in the growth rates. Nevertheless, this autocorrelation does not induce significant biases in our aggregate volatility calculations. We address the implications of autocorrelation in depth within Appendix C.

We next consider how to restrict the number of countries to exclude those least interested by French flows of exports. In this respect, a robust and consistent estimation of the destination-specific effects requires that factors have measurable impacts both at the macroeconomic and microeconomic levels. We keep in the dataset those countries that: (i) are sufficiently represented in the firms' portfolios; (ii) are relevant in terms of export value as a share of the total export.

Model dimensionality

After the filtering procedures, we are left with 67 destinations, accounting for 88.25% of the total export value. As we are ultimately interested in the growth rates, we further drop firm-country flows that over the whole span report data on three points or less (over 96). We are finally left with 86.44% of the total export value and close to 900,000 firm-destination pairs. To sum up, we estimate the dynamic factor models defined by equations (1)–(3) on a dataset composed by around 873,000 time series on time span comprising 96 quarters. On average, the information is condensed into the 25% of observations for which a growth rate is defined, while the remaining points are treated as missing values. The selected destinations shard the cross section into 67 blocks to which we associate as many local factors. Thus given the structure of the main model, which postulates the existence of a single global factor, we have 68 factors modelled as autoregressive processes of order one. As a result, in order to estimate the model we need to provide the estimates of the following static parameters: (i) 1746 thousands of factor loadings for each time series, one for the global factor and one for the related local factor, and (ii) 68 autoregressive coefficients from the definition of the dynamics of the factors.

Missing values distribution

As we can see from Figure, the operated spatial restriction induces a reduction in micro-level sparsity: the missing values represent around the 75% of the dataset. While the methodology as presented in Bańbura and Modugno (2014) does not require the missing values to follow a specific distribution for the model's identification and estimation, it is advised to look into the basic distributional properties for block factors applications. In the absence of a comprehensive theory detailing the effects of missing data on the convergence speed of the QML estimator, we turn to simulation experiments for guidance. Both the simulations from Bańbura and Modugno (2014) and our additional simulations provided in the Appendix B demonstrate that the process of estimating factors and loadings remains effective with up to 90% missing data.

In applying these insights to block factors, it is critical to monitor the relative amount of missing data within each block so that block-specific factors and loadings are consistently estimated. Figure 5 shows the results of an analysis of the distribution of missing data

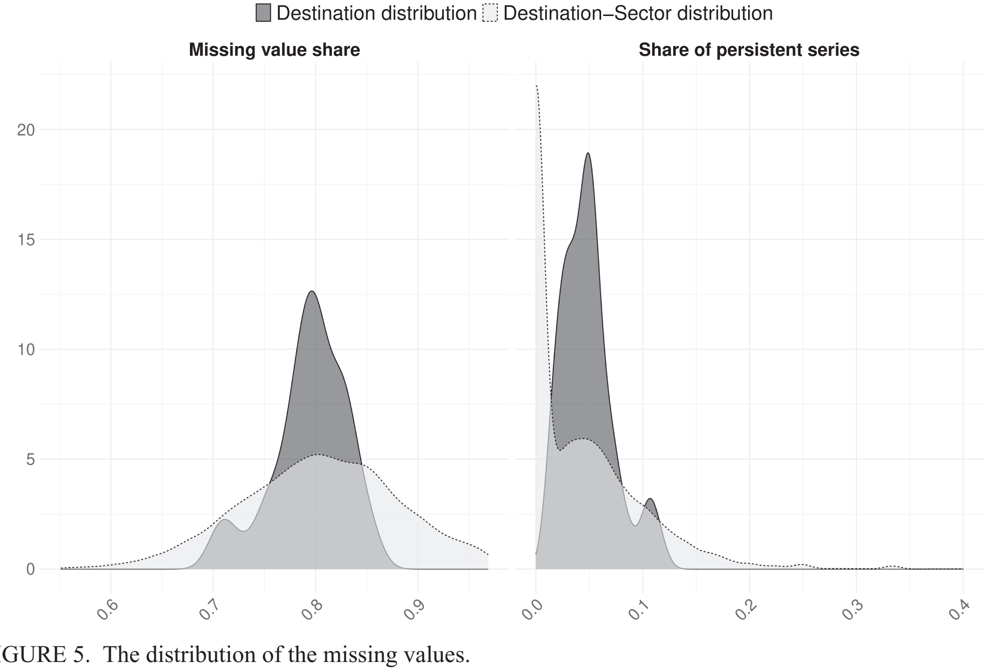


FIGURE 5. The distribution of the missing values.
Notes: Kernel density estimates of the average number of missing values and the number of persistent series across destinations (blocks).

using two separate block partitions: by destination and by a combination of destination and sector.[9] This analysis explores not only the average amount of missing data within these groups but also the prevalence of what we call 'persistent series' – those with at least 70% of available data points. Our analysis concludes with two key insights. The first is that the way we have organized our data into blocks is compatible with our chosen estimation method. This structure holds up well under the method's constraints as the right extreme of the destinations distribution reach at most the 87%. The figure also explains that a block structure obtained using both sectors and destinations poses challenges given the significant mass exceeding the 90% share threshold. The second insight is the notable presence of persistent series across all groups within our dataset, which we interpret as an indicator of the dataset's adequacy for capturing local co-movements. This also suggests that the distribution of missing data within blocks does not heavily concentrate towards the average, providing a more robust foundation for our analysis.[10]

IV. Results

In this section, we outline an overview of the main results of our analysis, grouped into two main categories. First, at the aggregate level, we show how dynamic factor models provide a robust identification of macroeconomic comovements that, together with firm-destination specific loadings, serve as the primary tool for the volatility decomposition. Using the decomposition, we show how the volatility associated with specific destinations distributes along geographical patterns typical of gravity models for export flows. Second, at the firm level, we look at the distributions of the components of firms' volatility and then characterize the linkages between geographical diversification and volatility trends.

Volatility at the aggregate level

Factor space reconstruction

Before moving ahead, we check the consistency of the factors' estimates from different samplings, controlling that the identified factor spaces are close enough. To this end, we compute the pairwise trace statistics with the formula[11]:

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

The range of the trace statistics is $[0, 1]$ and different factor spaces tend to be closer when $\mathrm{Tr}_{(k,h)}$ approaches the right limit. Trace statistics test for the equivalence of

[9] For this exercise, we assign each firm to the sector of main activity.

[10] Persistent series and average values are only synthetic indicators; a complete characterization of the distributions confirms these insights. The results of such an analysis are available upon request.

[11] F defines a $(D+1) \times T$-matrix containing both the global and the destination-specific factors. With a slight abuse of notation, we use the indices k and h to denote different samples of the dataset, not to be confused with the indices denoting the steps of the EM algorithm.

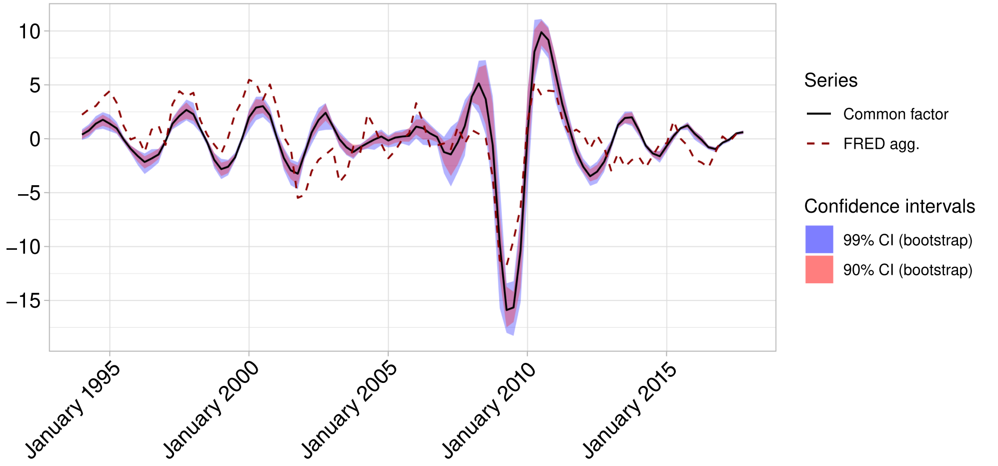


FIGURE 6. The global factor and the aggregate growth rate. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: The identified global factor (black solid line) with 90% and 99% confidence intervals, compared with the aggregate growth rate of the French export from an independent source: our elaboration of the series from Organization for Economic Co-operation and Development, Exports: Value Goods for France [XTEXVA01FRQ664N], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/XTEXVA01FRQ664N, 3 December 2020.

the factor spaces estimated by running the EM algorithm on each sample. The estimated matrices of factors from two different samples are compared by taking the related trace statistics. The pairwise computed values for 20 samples have a minimum of 0.96 and a maximum of 0.98, confirming a very good coherence of the different estimates.

The global factor is shown in Figure 6 where it is compared with yearly quarter-to-quarter growth rates for French exports from the FRED database. Simple visual inspection suggests a good level of agreement between the two independent measures of export growth, confirming that the comovements of the microeconomic export flows encode enough information to reconstruct the behaviour of aggregate statistics. We then extend the exercise on the global factor to see whether also the estimates of local factors admit economically meaningful interpretations by looking at the correlations with macroeconomic financial and economic indicators. Table 1 below and Table 7 in Appendix D show the results of a set of linear models and confirm the insights from Figure 6: the co-movements extracted from the microeconomic series correlate mainly with the bilateral export flows from independent sources, which contribute to more than the 70% of the variance explained by the full model. A significant impact for output growth differential and exchange rates (both nominal and real), with the former contributing to the large part of the remaining unexplained variance. Given the level of detail of our model, it is possible to test the correlation for each destination to study the diverse effects of the macroeconomic variables on the destinations' export. A full discussion is out of the scope of this paper, but the results of this exercise are presented in Appendix D.

TABLE 1

Correlation of local factors with country-specific financial and macroeconomic variables

	Destination factors		
	(1)	*(2)*	*(3)*
Bilateral flows GRs	0.026***	0.025***	0.025***
	(0.003)	(0.003)	(0.003)
Int. rate diff.		0.005	0.007
		(0.014)	(0.014)
Output growth diff.		0.046***	0.047***
		(0.009)	(0.009)
Nom. ex. rate		0.0001	0.0001
		(0.001)	(0.001)
Real ex. rate		0.456***	0.444***
		(0.151)	(0.150)
Nom. ex. rate GRs			0.002***
			(0.001)
Real ex. rate GRs			0.0003*
			(0.0002)
Intercept	0.0002	−0.483***	−0.463***
	(0.065)	(0.179)	(0.176)
Destination FE	Yes	Yes	Yes
Observations	6,336	2,149	2,149
R^2	0.027	0.075	0.079
Adjusted R^2	0.017	0.062	0.066
Residual SE	0.759 (df = 6269)	0.643 (df = 2119)	0.642 (df = 2117)
F Statistic	2.678*** (df = 66; 6269)	5.898*** (df = 29; 2119)	5.897*** (df = 31; 2117)

Notes: Relative importance of the regressors in (3). Heteroscedasticity-robust SEs in parantheses. Bilateral Flows GRs: 72.84%, Out. Growth Diff.: 19.20%, Nom. Ex. Rate GRs: 3.53%, Other vars (cumulative): 4.43%. Exchange rates and interest rates variables come from the International Financial Statistics (IFS) section of the International Monetary Fund (IMF). Bilateral trade flows are available at the Direction of Trade Statistics of the IMF. Output Growth are collected from the Quarterly National Accounts of the OECD. See Appendix D for details on the variables definition and the transformations applied.

$^{*}P < 0.1$; $^{**}P < 0.05$; $^{***}P < 0.01$.

Aggregate volatility and granularity

Before providing the point estimates of the aggregate volatility components, we present a comparison between the growth rates decomposition obtained via SODMs and DFMs. As a benchmark for the former we replicate on our dataset the decomposition methodology of di Giovanni *et al.* (2014).[12] For a proper comparison, we analyse the common term of our decomposition, joining the global and local factor and loadings (i.e. the firs two terms of equation 1), vis a vis the sector-destination shocks of di Giovanni *et al.* (2014) (the first term in equation 5, pag. 1309).[13] Looking at Figure 7 we see in which direction our decomposition enriches the analysis of the variation in the export sales growth rates. Both estimates are in line with the previous findings highlighting that the variation induced by the idiosyncratic shocks is dominant in magnitude over the common shocks for most of the time span. Nevertheless, it is worth emphasizing two relevant differences: (i) the common components derived as the interplay between factors and the related firm responses are (i)

[12] The algorithms are adapted from those available at the link https://julian.digiovanni.ca/Papers/FirmGranular_replication.zip.

[13] In the following we will always specify if the calculations include both sectors and destinations effect or only one of the two.

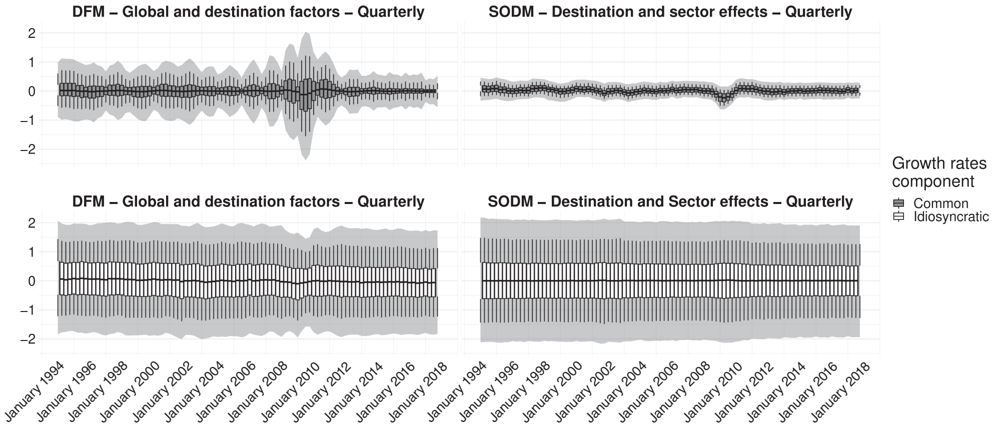
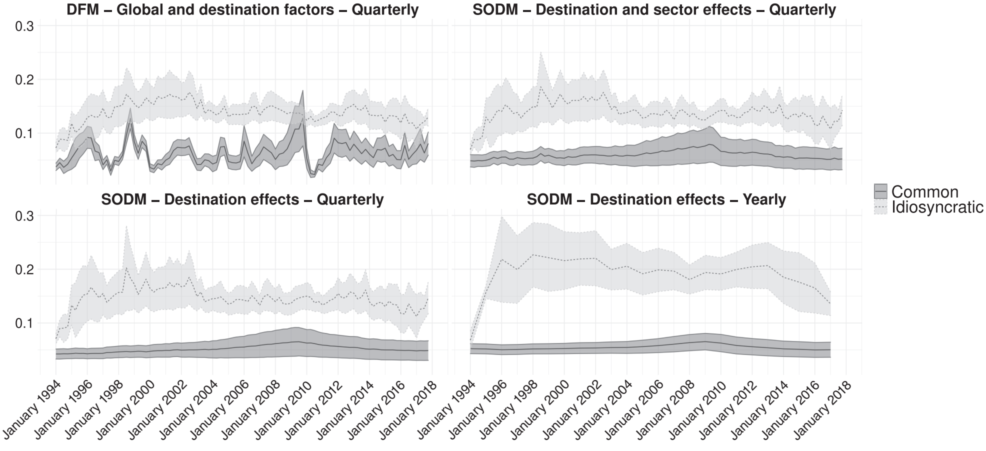


FIGURE 7. Growth rates components distribution for SODMs and DFMs estimates.
Notes: The direct comparison of the growth rates common (top) and idiosyncratic (bottom) components obtained via dynamic factor models (top left) or the static orthogonal decomposition models of di Giovanni *et al.* (2014) under different specifications, varying frequency and macroeconomic effects. The distributions elements shown are the components' median (box centers), the first and third quartiles (box extremes) and the first and last deciles (segment extremes), the 5th and 95th percentiles (background shadowed area)

significantly more volatile than common components extracted as sectoral and destination averages; (ii) this evidence is even more relevant across specific subsets of the time span, with some cases in which the common component isolated through DFMs approaches the idiosyncratic shocks in magnitude. These results highlight the flexibility of DFMs in capturing non-trivial effects of the macroeconomic components on the export growth rates at microeconomic level. This evidence leads to a partial review of the standard results that most shocks hitting firms are firm-specific, suggesting instead that also shocks common to all the firms can generate significant variations in the growth rates because firms' reactions are highly heterogeneous. Not surprisingly, such a mechanism becomes particularly relevant during deep downturns and rapid hypes.

After a first exploration of the decompositions of the growth rates, we move to the analysis of the aggregate volatility to assess the differential impact of the decompositions. Using equation (6) the aggregate volatility estimates are determined not only by the variation in growth rates components but also by the possible synchronization with the changes in the distribution of the size-based weights. In this respect, Figure 8 and Table 2 provide the aggregate volatility estimates respectively at each time step and considering all the time span. The firm-destination specific component accounts for the 0.84 of the actual volatility; the global and destination-specific volatility components for the 0.30 and the 0.18, respectively. When both the business cycle terms are combined to form the common component they reach a relative share of 0.37 over the time span, in line with the outcomes of the SODM, setting around the 0.36 and 0.31 depending if both sector and destination effects are included or destination effects only. As before, the comparison shows a significant difference when we zoom into the details of the time variations. In fact, in our model the dynamic of the volatility is not only driven by the variation of the relative size of the exporters or the exporter-destination cells (as measured by the weights), as in di Giovanni *et al.* (2014), but also by the significant variation in the growth rates and in particular by their synchronization. This might induce relevant changes in the estimates



Time distribution of the ratio between the common and idiosyncratic components

	Min	q10	q50	q90	Max
DFM (Dest.)	0.1886	0.2869	0.4524	0.6443	0.9566
Period	2010–04	2006–07	2002–01	1995–04	2009–10
SODM (Dest. + Sec.)	0.3138	0.3554	0.4197	0.5457	0.6967
Period	1999–10	1996–07	2003–10	1994–04	2009–10

FIGURE 8. Volatility components estimates for SODMs and DFMs.
Notes: The direct comparison of the aggregate volatility estimates obtained decomposing the growth rates via DFMs or the SODMs. The latter is presented under two different specifications, namely including destination and sector terms in the common component, or destination effects only. The SODM is estimated both on quarterly and yearly data to highlight the impact on the estimates of a change in the measured frequency. Shadowed areas outline the 95% analytical confidence intervals as provided in di Giovanni *et al.* (2014). The table shows selected quantiles of the time-distribution (at a quarterly frequency) of the ratio between common and idiosyncratic components.

TABLE 2
Volatility components as a share of the aggregate volatility

	DFM (sampled)	*SODM*			
	Dest.	*Dest.*		*Dest. + Sec.*	
	(Quarterly)	*(Quarterly)*	*(Yearly)*	*(Quarterly)*	*(Yearly)*
Constant weighted aggregation					
Common	0.3753 (0.3579,0.3913)	0.3173	0.2608	0.3630	0.2913
Global	0.3049 (0.2594,0.3405)				
Destination	0.1800 (0.1539,0.2063)				
Idiosyncratic	0.8444 (0.8369,0.8506)	0.8915	0.9073	0.8498	0.8722
Dynamic weighted aggregation					
Common	0.7853 (0.7431,0.8188)	0.6484	0.5414	0.7171	0.5786
Global	0.684 (0.5992,0.7651)				
Destination	0.3683 (0.3150,0.4399)				
Idiosyncratic	0.7873 (0.7653,0.8106)	0.7904	0.5367	0.7235	0.5108

Notes: The volatility explained by each component in relation to the actual aggregate volatility as measured by the DFMs and the benchmark SODMs. The statistics and confidence intervals for the DFM are computed out of 20 estimations of the same model over random subsamples of the original dataset (each subsample is constructed selecting 80k firms at random).

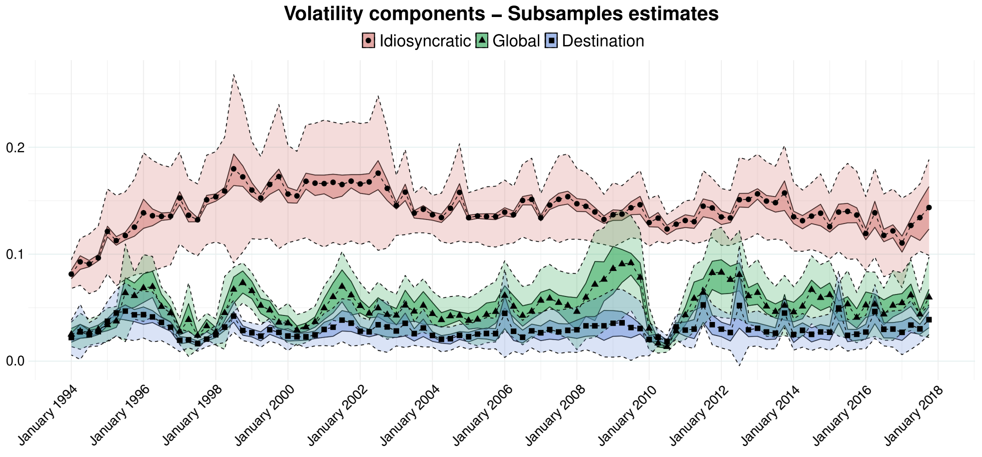


FIGURE 9. Volatility components. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: The estimates of the aggregate volatility components, global, destination-specific and idiosyncratic components obtained via dynamic factor models from subsample estimates (average of the point estimates), aggregating with constant weights. Confidence bands: darker with solid line, 5th and 95th percentile of the point volatility estimates from different samples; lighter with dashed line, we take the maximum across subsamples of the 95% bootstrapped confidence interval derived as in di Giovanni *et al.* (2014).

of the aggregate volatility as the time span changes: note, for example, how the common component of the volatility surged during the trade collapse matching in magnitude the idiosyncratic one. These differences do not disappear if we change the time-frequency or the set of effects that are included in the benchmark decomposition. The table at the bottom of Figure 8 shows how the time-distribution of the ratio between the common and idiosyncratic components can be very different: the SODM with the highest number of fixed effects underestimates the variation range of the common volatility over time. At its maximum, the common component almost matches the idiosyncratic one when estimated via DFMs (ratio at 95.66%), while set just below 70% for the SODMs estimates.

We conclude this part, with the presentation of the results from the subsampling exercise. Figure 9 reports the time series of the volatility estimates for all the components. The common component is exploded into its subcomponents, global and destination-specific, and presented together with the idiosyncratic one. Destination and global components are comparable in magnitudes except for a few spikes of the global component standing out during downturns of the cycles. The idiosyncratic component is dominant in magnitude, as mentioned before. This pattern is reflected in the point estimates of the volatility associated to the whole time-span in Table 2, top panel.

Destination-specific patterns

In this subsection, we further investigate the structure and characteristics of the destination-specific components associated with aggregate French exports. To guide our analysis, we start from the gravity hypothesis, which is usually concerned with the cross-section of trade flows. However, we expect that ‘gravity-like patterns’ are also relevant for understanding the underlying dynamics of trade volatility. Specifically, we argue that the microstructure of bilateral trade flows, influenced by gravity-like factors, directly affects the temporal volatility of export growth. The literature highlights how distance – interpreted in various

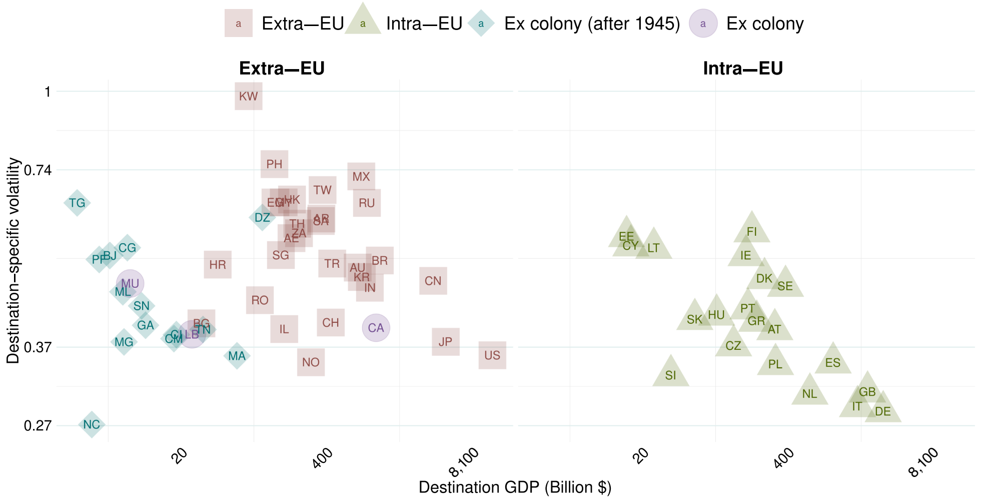


FIGURE 10. Destination-specific volatility vs destination GDP. [Colour figure can be viewed at *wileyonlinelibrary.com*]
Notes: Point diagram for extra-EU (left panel) and intra-EU (right panel) trade relations.

ways (Mehl *et al.*, 2019) – amplifies the effects of uncertainty due to financial shocks (Berman *et al.*, 2013) or simple inventory management (Bekes *et al.* 2017). Consistent with these findings, our results indicate a diversification effect within destinations, which helps mitigate the amplification of uncertainty, particularly as the number of firms engaged in trade transactions increases and distances decrease, or when dealing with larger economies (Bernard *et al.*, 2007). Figure 10 shows a simple bivariate relation between the measure of volatility associated to a given destination and its GDP level. In the right panel, where the analysis is restricted to EU countries, a clear inverse correlation emerges between volatility and GDP. Indeed, the outliers, if any, are EU commercial partners joining the European Union in the second half of our time window.[14]

Similar considerations apply to the patterns on the left panel: an inverse relationship is apparent among the Extra-EU countries, which is however obscured by the presence of specific group of countries which are likely to experience less volatile trade flows from French firms, for a given level of GDP. This is the case, for example, of former French colonies status, which reduce the risks associated to trade relationships with low income countries.

To provide a more quantitative assessment of these trends, we ran a simple OLS regression of destination specific volatility on GDP, further adding controls accounting for geographical distance, free trade agreements, former French colonies, EU countries. Results, presented in Table 3, confirm the main finding from Figure 10: when accounting for additional covariates, the relationship between (log) GDP and (log) destination-specific volatility becomes negative and significant (see columns 2 and 3).

[14]Slovakia and Lithuania joined the EU single market in 2004, while the analysed time window spans from 1993 to 2017.

TABLE 3

Gravity-like linear models for the destination-specific volatility

	Log. dest. vol.		
	(1)	*(2)*	*(3)*
Log. dist.	0.123***	0.096***	0.097***
	(0.029)	(0.034)	(0.031)
Log. GDP	0.004	−0.053***	−0.040**
	(0.014)	(0.017)	(0.017)
EU+Colonies controls	No	Yes	No
FTA+Colonies controls	No	No	Yes
Constant	−1.994***	−0.172	−0.513
	(0.444)	(0.553)	(0.486)
Observations	65	65	65
R^2	0.231	0.460	0.501
Adjusted R^2	0.206	0.415	0.430
Residual SE	0.218 (df = 62)	0.187 (df = 59)	0.185 (df = 56)
F Statistic	9.309*** (df = 2; 62)	10.064*** (df = 5; 59)	7.029*** (df = 8; 56)
Note:			$*P < 0.1$; $**P < 0.05$; $^{***}P < 0.01$

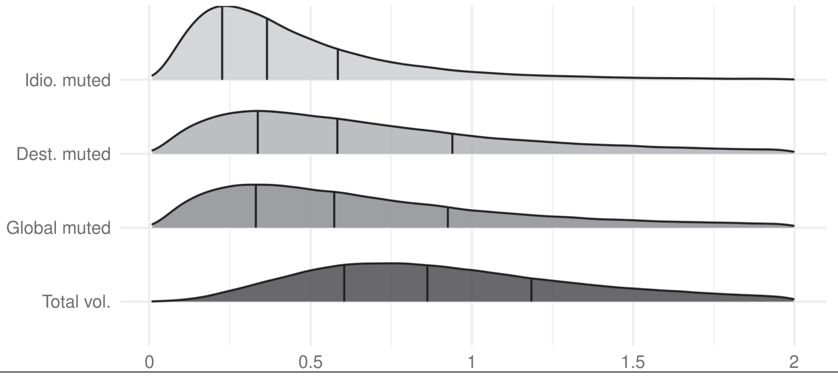
Notes: OLS estimates of gravity-like regressions for the volatility associated to each destination. GDP values and distances are the variables *distw* and *gdp_d* with France taken as origin country from the CEPII GeoDist database (available at http://www.cepii.fr/CEPII/en/bdd_modele/presentation.asp?id6 and described in Head *et al.*, 2010; Mayer and Zignago, 2011). EU+Col controls include the variables *eu_d*, *col45* and, *colony*, whereas for FTA+Col *eu_d* is replaced by *fta_bb*; the latter includes information on the participation in a common market or in the same currency union.

Volatility at the firm level

In order to establish the role of the different components of volatility at the firm level, Figure 11 shows the distributions of firm-level volatility where one or two terms of equation (4) have been set to zero before aggregating using the dynamic weights of the form (9), in the spirit of Kramarz *et al.* (2020). The global and destination-specific components have similar impacts on firms' volatility distribution, both in terms of magnitude and direction: once muted singularly, there is a visible shift to the left of the second and third quartile threshold and, to a lesser extent, 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 all quartile thresholds and a substantial narrowing of the right tail. More precisely, the median reduction obtained removing the macroeconomic components is around 50% (from 0.78 to 0.37 or 0.39). In contrast, the impact of the microeconomic (idiosyncratic) contribution amounts to a dampening of 83% (from 0.78 to 0.13), confirming the prominent role of the idiosyncratic component, yet showing that global and destination-specific terms have a non-negligible effect as drivers of volatility. This means that exporters, even though mostly exposed to idiosyncratic risks, face also the risk of relevant global and destination-specific shocks. In the remainder, the focus will move on the role of diversification in mitigating the effects of these shocks.

Measuring diversification

To investigate the relationship between export growth volatility and firm diversification, we start by defining a set of diversification measures that have been used in the literature: the number of destination markets (Dest. Mkts.), the share of firm exports accounted for



	10th pct.	25th pct.	50th pct	75th pct.	90th pct.
Total vol.	0.1754	0.3723	0.7835	1.5758	2.9223
Glo. muted	0.0351 (−80%)	0.1149 (−69%)	0.3714 (−53%)	1.1051 (−29%)	3.0645 (+5%)
Dest. muted	0.0371 (−79%)	0.1198 (−68%)	0.3881 (−50%)	1.1588 (−26%)	3.2063 (+ 9%)
Idio. muted	0.0182 (−89%)	0.0478 (−87%)	0.1301 (−83%)	0.3497 (−78%)	0.8725 (−70%)

FIGURE 11. Kernel density estimates for the ‘counterfactual’ volatility distributions.
Notes: On each panel a fit of the empirical distributions is obtained silencing the idiosyncratic component (top panel) and the macroeconomic effects: destination-specific component and global component (middle panels). The table below provides a quantitative assessment of the shift of the quantile thresholds for the same distributions.

TABLE 4
Distribution of the diversification indicators

	Number of obs.	*Mean*	*SD*	*Skew.*	*Kurt.*	*Median*	*1st pc.*	*10th pc.*	*90th pc.*	*99th pc.*
Dest. Mkts.	143,194	2.12	4.98	6.89	76.52	0.58	0.07	0.13	5.09	24.21
Top Share	143,194	0.28	0.18	1.01	0.54	0.24	0.05	0.09	0.55	0.81
HH Ind.	143,194	0.74	0.17	−1.16	1.21	0.79	0.23	0.50	0.92	0.96
Div. Ind.	143,194	1.69	1.30	3.43	16.75	1.12	1.00	1.00	3.14	7.25

Notes: Summary statistics for the distributions of four key diversification indicators: the number of destination markets, the share (of value) of the principal destination market in the firm’s portfolio, the Herfindahl−Hirschman Index (HHI) and a diversification index constructed using the inverse of the HHI. The synthetic information is computed as an average over the active time span for each single exporter.

by the most important market (Top Share), and the Herfindahl-Hirschman Index (HHI) of export shares (see, among many others, Braakmann and Wagner, 2011; di Giovanni *et al.*, 2014; Vannoorenberghe *et al.*, 2016; Kramarz *et al.*, 2020). We add to this indices also a diversification measure constructed using the inverse of the HHI. For each of these variables, we compute the firm average over time on a quarterly basis and then report basic descriptive statistics (Table 4). We observe a consistent level of skewness in all the relevant distributions, in line with the previous evidence of Eaton *et al.* (2004) on French exporters.

Volatility components and diversification

The preceding analysis has one main implication: the idiosyncratic component of the firm-level volatility has a prominent role compared to the macroeconomic ones, i.e. the global component and the destination-specific component. How do diversification strategies help firms reduce trade risks related to the different components? Start by

TABLE 5
Correlation matrix of the diversification indicators

	σ	*Dest. mkts.*	*Top share*	*HH ind.*
Dest. mkts.	−0.1627			
Top Share	−0.0893	0.1592		
HH Ind.	0.0489	−0.0292	−0.9824	
Div. Ind.	−0.2230	0.8022	0.0072	0.1268

Note: Pearson correlations between different diversification measures and log weighted volatility.

noting that standard portfolio theory would imply a negative relationship between firm-level volatility and the degree of diversification in the destination markets (Hirsch and Lev, 1971; Vannoorenberghe *et al.*, 2016). Table 5, showing pairwise correlations among the different diversification indicators and firm-level volatility, provides a descriptive confirmation that more diversified firms tend indeed to experience less volatile export growth patterns.

To dig deeper into these correlations, we look at the distributional properties of the volatility components for classes of firms grouped by diversification quantiles, in Figures 12 (for all firms) and 13 (focusing only on high and low volatile firms). First, the macroeconomic volatility components move together with a downward trend in logs. Second, the idiosyncratic component moves along two opposite trends if we look at exporters that diversify below or above the median. Indeed, looking at the first half of the diversification spectrum, the volatility lies on a steady path, 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 relatively high for the population of French exporters, corresponding to a diversification index of around 3. On our data, the median value of the diversification index is 1.12, and only the 10% most diversified firms reach that limit (See Table 4, bottom). Figure 13, which focuses only on the most and the least volatile firms, additionally confirms that the idiosyncratic component dwarfs the other two in magnitude.

One concern raised by Figures 12 and 13 is that they do not control for size effects: as more diversified firms are also larger, the negative relationship between diversification and volatility could be due to an underlying size effect. To account for this, Figure 14 shows the relationship between volatility and its components and the residuals from a OLS regression of each of these components on size percentiles. The main message does not change: there is a sharp negative relationship between diversification and the common and destination-specific components of volatility. On the other hand, the idiosyncratic component is much less responsive to diversification.

Summing up, the risk exposure is reduced for firms that diversify their activities on the destination markets. Log-linear risk dampening effects seem to work only for shocks originated by macroeconomic-induced fluctuations, 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. Beyond the threshold, diversification strategies give a consistent reduction helping firms to approach less volatile growth paths.

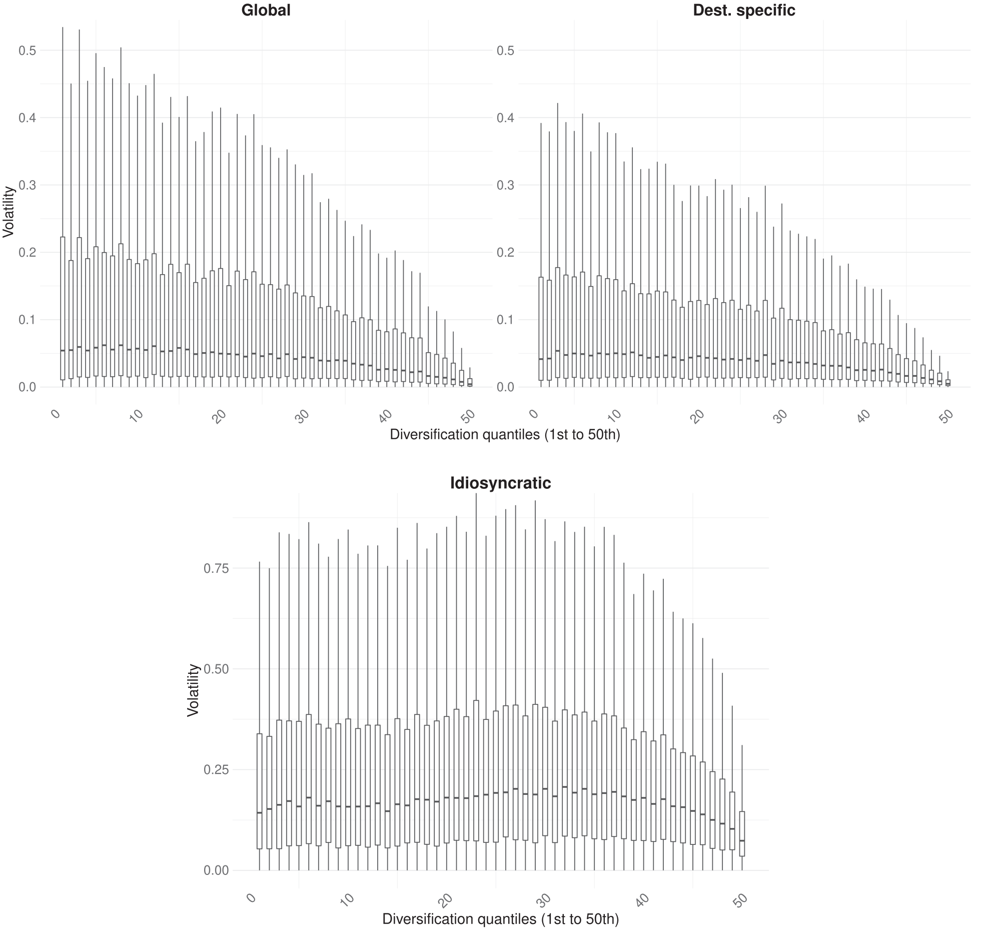


FIGURE 12. Diversification vs. volatility components
Notes: The population of firms is divided into 50 groups, one for each 50tile of diversification indicator (Div. Ind. as per Table 4). The graphs display the boxplot associated with the distribution of the volatility components of each group, ordered from the least to the most diversified.

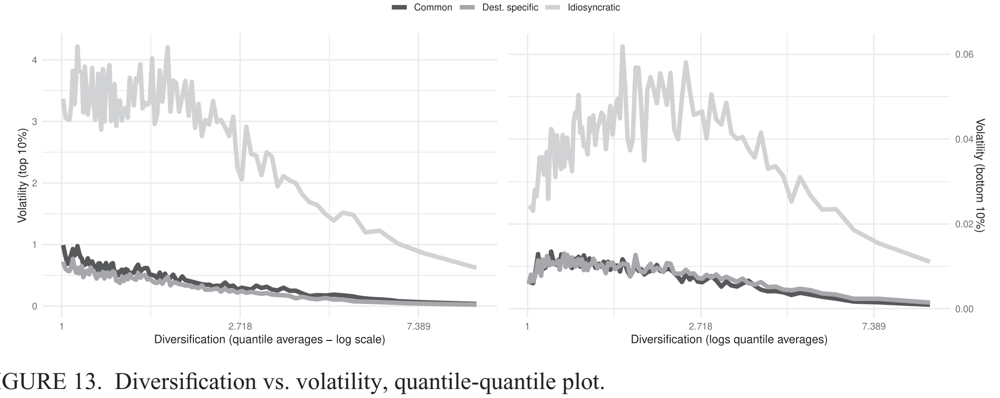


FIGURE 13. Diversification vs. volatility, quantile-quantile plot.
Notes: On the left panel, restricted to the most volatile firms (top 10%), the plot shows the three different volatility components vs percentiles of inverse Herfindahl index (average on the time window). On the right panel, the analogous graph for the least volatile firms (bottom 10%).

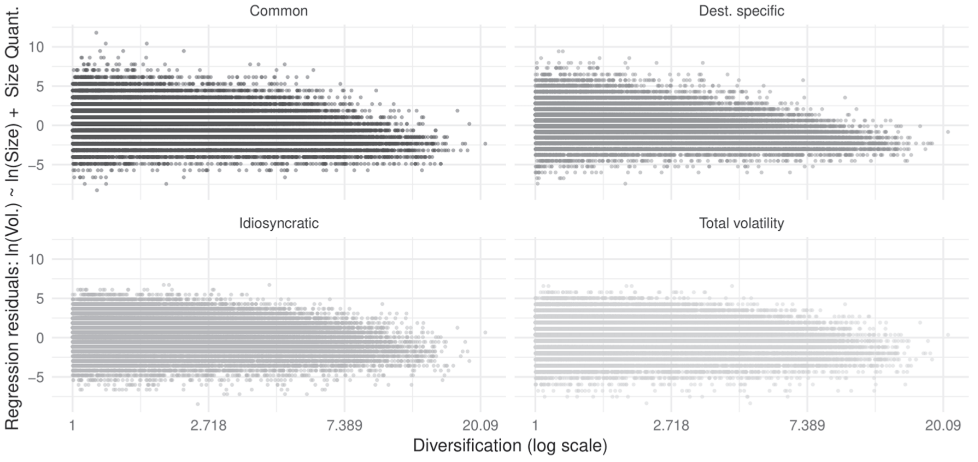


FIGURE 14. Diversification vs. volatility controlling for size effects.
Notes: The residuals of the regression ln(Vol) $\sim$ ln(Size) + Size percentiles plotted against the diversification index. As explained variable we take each component of the volatility and the volatility itself at the firm-level.

V. Conclusions

This paper proposes a dynamic factor model approach to the decomposition of aggregate and firm-level volatility. This allows to reconstruct the latent space of macroeconomic factors and decompose the growth rate of firm-destination cells into three orthogonal components: a global component, a destination-specific component and an idiosyncratic component. This provides the first application of dynamic factor models to transaction level data and requires an estimation based on the EM algorithm to accommodate both the prevalence of missing values and the high number of time series.

The decomposition is then mapped at the aggregate and at the firm level measuring the contribution of the three components to the total export's and firms' volatility. Our method gives new insights on the impact of the granular component of the aggregate export volatility and its measurement. In particular, we find that macroeconomic shocks play a bigger role in generating aggregate volatility than it is usually recognized.

When analyzing the volatility associated with firms' growth, we show that the global and destination-specific components have a comparable effect on the first and second moments of the volatility distribution and show how diversification across destination markets can protect firms from shocks coming from macroeconomic events but seems to have contrasting effects on the idiosyncratic shocks.

ACKNOWLEDGMENT

Open access publishing facilitated by Universita degli Studi di Bologna, as part of the Wiley - CRUI-CARE agreement.

Final Manuscript Received: March 2023

References

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

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

Armenter, R. and Koren, M. (2014). 'A balls-and-bins model of trade', *American Economic Review*, Vol. 104, pp. 2127–2151.

Bai, J. (2003). 'Inferential theory for factor models of large dimensions', *Econometrica*, Vol. 71, pp. 135–171.

Bai, J. and Li, K. (2016). 'Maximum likelihood estimation and inference for approximate factor models of high dimension', *Review of Economics and Statistics*, Vol. 98, pp. 298–309.

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

Barigozzi, M. and Luciani, M. (2019). *Quasi Maximum Likelihood Estimation and Inference of Large Approximate Dynamic Factor Models via the EM Algorithm*. arXiv preprint arXiv:1910.03821.

Bekes, G., Fontagné, L., Muraközy, B. and Vicard, V. (2017). 'Shipment frequency of exporters and demand uncertainty', *Review of World Economics*, Vol. 153, pp. 779–807.

Bergounhon, F., Lenoir, C. and Mejean, I. (2018). *A Guideline to French Firm-Level Trade Data*.

Berman, N., De Sousa, J., Martin, P. and Mayer, T. (2013). 'Time to ship during financial crises', in *NBER International Seminar on Macroeconomics* Vol. 9, University of Chicago Press, Chicago, IL, pp. 225–260.

Bernard, A. B., Boler, E. A., Massari, R., Reyes, J.-D. and Taglioni, D. (2017). 'Exporter dynamics and partial-year effects', *American Economic Review*, Vol. 107, pp. 3211–3228.

Bernard, A. B., Jensen, J. B., Redding, S. J. and Schott, P. K. (2007). 'Firms in international trade', *Journal of Economic Perspectives*, Vol. 21, pp. 105–130.

Bernard, A. B., Jensen, J. B., Redding, S. J. and Schott, P. K. (2009). 'The margins of US trade', *American Economic Review*, Vol. 99, pp. 487–493.

Bernard, A. B., Jensen, J. B., Redding, S. J. and Schott, P. K. (2012). 'The empirics of firm heterogeneity and international trade', *Annual Review of Economics*, Vol. 4, pp. 283–313.

Bernard, A. B., Jensen, J. B., Redding, S. J. and Schott, P. K. (2016). *Global Firms*, CEP Discussion Papers No. dp1420, Centre for Economic Performance, LSE.

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

Breitung, J. and Eickmeier, S. (2014). *Analyzing Business and Financial Cycles Using Multi-Level Factor Models*. CAMA Working Paper No. 43/2014. http://dx.doi.org/10.2139/ssrn.2443639

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

Bricongne, J.-C., Carluccio, J., Fontagné, L. G., Gaulier, G. and Stumpner, S. (2022). *From Macro to Micro: Large Exporters Coping with Common Shocks*, CESifo Working Paper No. 9857.

Carvalho, V. and Gabaix, X. (2013). 'The great diversification and its undoing', *American Economic Review*, Vol. 103, pp. 1697–1727.

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

Choi, I., Kim, D., Kim, Y. J. and Kwark, N.-S. (2018). 'A multilevel factor model: Identification, asymptotic theory and applications', *Journal of Applied Econometrics*, Vol. 33, pp. 355–377.

Coroneo, L., Giannone, D. and Modugno, M. (2016). 'Unspanned macroeconomic factors in the yield curve', *Journal of Business & Economic Statistics*, Vol. 34, pp. 472–485.

Delle Chiaie, S., Ferrara, L. and Giannone, D. (2022). 'Common factors of commodity prices', *Journal of Applied Econometrics*, Vol. 37, pp. 461–476.

di Giovanni, J. and Levchenko, A. A. (2009). 'Trade openness and volatility', *The Review of Economics and Statistics*, Vol. 91, pp. 558–585.

di Giovanni, J., Levchenko, A. A. and Mejean, I. (2014). 'Firms, destinations, and aggregate fluctuations', *Econometrica*, Vol. 82, pp. 1303–1340.

Di Giovanni, J., Levchenko, A. A. and Mejean, I. (2018). 'The micro origins of international business-cycle co-movement', *American Economic Review*, Vol. 108, pp. 82–108.

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

Eaton, J., Kortum, S. and Kramarz, F. (2004). 'Dissecting trade: firms, industries, and export destinations', *American Economic Review*, Vol. 94, pp. 150–154.

Freyaldenhoven, S. (2022). 'Factor models with local factors — determining the number of relevant factors', *Journal of Econometrics*, Vol. 229, pp. 80–102.

Gabaix, X. (2011). 'The granular origins of aggregate fluctuations', *Econometrica*, Vol. 79, pp. 733–772.

Hallin, M. and Liška, R. (2011). 'Dynamic factors in the presence of blocks', *Journal of Econometrics*, Vol. 163, pp. 29–41.

Head, K., Mayer, T. and Ries, J. (2010). 'The erosion of colonial trade linkages after independence', *Journal of international Economics*, Vol. 81, pp. 1–14.

Herskovic, B., Kelly, B., Lustig, H. and Van Nieuwerburgh, S. (2020). 'Firm volatility in granular networks', *Journal of Political Economy*, Vol. 128, pp. 4097–4162.

Hirsch, S. and Lev, B. (1971). 'Sales stabilization through export diversification', *The Review of Economics and Statistics*, Vol. 53, pp. 270–277.

Kelly, B., Lustig, H. and Nieuwerburgh, S. V. (2013). *Firm Volatility in Granular Networks*, NBER Working Papers No. 19466, National Bureau of Economic Research, Inc.

Kose, M. A., Otrok, C. and Prasad, E. (2012). 'Global business cycles: convergence or decoupling?', *International economic review*, Vol. 53, pp. 511–538.

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

Marcellino, M. and Schumacher, C. (2010). 'Factor MIDAS for nowcasting and forecasting with ragged-edge data: a model comparison for German GDP', *Oxford Bulletin of Economics and Statistics*, Vol. 72, pp. 518–550.

Mariano, R. S. and Murasawa, Y. (2003). 'A new coincident index of business cycles based on monthly and quarterly series', *Journal of applied Econometrics*, Vol. 18, pp. 427–443.

Mayer, T. and Zignago, S. (2011). *Notes on CEPII's Distances Measures: The GeoDist Database*. CEPII Working Paper 2011-25.

Mehl, A., Schmitz, M. and Tille, C. (2019). *Distance(s) and the Volatility of International Trade(s)*. ECB Working Paper Series No. 2252.

Mésonnier, J.-S. and Stevanovic, D. (2017). 'The macroeconomic effects of shocks to large banks' capital', *Oxford Bulletin of Economics and Statistics*, Vol. 79, pp. 546–569.

Miranda-Agrippino, S. and Rey, H. (2020). 'US monetary policy and the global financial cycle', *The Review of Economic Studies*, Vol. 87, pp. 2754–2776.

Moench, E. and Ng, S. (2011). 'A hierarchical factor analysis of US housing market dynamics', *The Econometrics Journal*, Vol. 14, pp. C1–C24.

Moench, E., Ng, S. and Potter, S. (2013). 'Dynamic hierarchical factor models', *Review of Economics and Statistics*, Vol. 95, pp. 1811–1817.

Mumtaz, H. and Surico, P. (2012). 'Evolving international inflation dynamics: world and country-specific factors', *Journal of the European Economic Association*, Vol. 10, pp. 716–734.

Onatski, A. (2012). 'Asymptotics of the principal components estimator of large factor models with weakly influential factors', *Journal of Econometrics*, Vol. 168, pp. 244–258.

Quah, D. and Sargent, T. J. (1993). 'A Dynamic Index Model for Large Cross Sections', In Stock J. H., and Watson M. W. (eds), *Business Cycles, Indicators, and Forecasting*, Studies in Business Cycles, Vol. 28, Univ. Chicago Press (for NBER), Chicago.

Reis, R. and Watson, M. W. (2010). 'Relative goods' prices, pure inflation, and the Phillips correlation', *American Economic Journal: Macroeconomics*, Vol. 2, pp. 128–157.

Sargent, T. J. and Sims, C. A. (1977). 'Business cycle modeling without pretending to have too much a priori economic theory', in *New Methods in Business Cycle Research* Vol. 1, Federal Reserve Bank of Minneapolis, Minneapolis, pp. 145–168.

Shumway, R. H. and Stoffer, D. S. (1982). 'An approach to time series smoothing and forecasting using the EM algorithm', *Journal of Time Series Analysis*, Vol. 3, pp. 253–264.

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

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

Watson, M. W. and Engle, R. F. (1983). 'Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models', *Journal of Econometrics*, Vol. 23, pp. 385–400.

Supporting Information

Additional Supporting Information may be found in the online Appendix:

Data S1. Online appendices.

Data replication package: the data replication package is available at https://doi.org/10.3886/E188961