NIPE Núcleo de Investigação em Políticas Económicas e Empresariais

#08
2020

WORKING PAPER

Robert F. Engle
Susana Campos-Martins

"Measuring and Hedging Geopolitical Risk"

https://www.eeg.uminho.pt/pt/investigar/nipe

FCT Fundação para a Ciência e a Tecnologia

«This paper is financed by National Funds of the FCT – Portuguese Foundation for Science and Technology within the project «UIDB/03182/2020»

Universidade do Minho
Escola de Economia e Gestão

Measuring and Hedging Geopolitical Risk

Robert F. Engle

New York Stern School of Business

Susana Campos-Martins

Nuffield College, University of Oxford

Centre for Research in Economics and Management,

University of Minho

Abstract

Geopolitical events can impact volatilities of all assets, asset classes, sectors and countries. It is shown that innovations to volatilities are correlated across assets and therefore can be used to measure and hedge geopolitical risk. We introduce a definition of geopolitical risk which is based on volatility shocks to a wide range of financial market prices. To measure geopolitical risk, we propose a statistical model for the magnitude of the common volatility shocks. Accordingly, a test and estimation methods are developed and studied using both empirical and simulated data. We provide a novel explanation for why idiosyncratic volatilities comove based on a new way to formulate multiplicative factors. Finally, we propose a new criterion for portfolio optimality which is intended to reduce the exposure to geopolitical risk.

1 Introduction

Geopolitical risk has become an increasingly important component of risk analysis. It is broadly defined as the exposure of one or more countries to political actions in other countries. Clearly events such as the *Brexit* referendum in 2016 are considered geopolitical events. However many other events such as military or terrorist actions and central bank or regulatory actions can also be interpreted as geopolitical events. Even local financial events, cyber attacks, trade wars and climate change can have global financial impacts.

In this paper we develop an empirical measure of geopolitical risk by defining it as a common shock to the volatility of a very wide class of financial assets. Geopolitical events are assumed to affect all countries, all asset classes, and all sectors. We will use the term *GEOVOL* to refer to such shocks. These shocks can be described as political,

regulatory, military, terrorist or natural disasters, but the key feature is that they move financial prices of a very wide class of assets.

To measure geopolitical risk, we use financial market prices which are assumed to incorporate all available information. A statistical approach to estimation is introduced and examined theoretically, by simulation and by data analysis. The results are then compared with other estimates which use different methodologies. Other major contributions are derived from the GEOVOL model. It provides a novel explanation for why idiosyncratic volatilities comove. It is also a new way to formulate multiplicative factors for volatility rather than the more traditional additive decomposition.

It is well known that volatilities of asset returns comove. It is natural to observe common variation when assets are all exposed to the same factors. If financial returns are linear combinations of common factors, time-varying factors will imply a volatility factor structure. However, whatever factors are taken out of the data, the idiosyncratic returns still have correlated volatilities (Herskovic et al., 2016). Since volatility is also well known to be predictable, the comovements of volatilities are most likely caused by correlation between the shocks to volatility. The fundamental observation underlying the GEOVOL model is that even though the volatility standardized residuals are orthogonal in both times series and cross section with unit variances, their squares can be correlated. This observation in time series was the key motivation for the original ARCH model of Engle (1982) and is now the key motivation for the GEOVOL model in cross section. It recognizes the fact that assets in different asset classes, sectors, or countries all respond to geopolitical news. Results thus extend the literature on idiosyncratic volatility including Connor et al. (2006) and Ang et al. (2006).

Comovements of innovations to volatilities are likely to be the primary source of the comovements of volatilities. It is thus natural to detect volatility factors by observing cross-sectional positive correlations between squared standardized innovations. This new and testable observation is first presented in the baseline model as the motivation to modeling and testing GEOVOL effects. The impacts of this common volatility factor on the volatilities of the asset returns, despite being at the same time, may be different across assets. To model this heterogeneity, the model is extended to allow for different factors loadings on the GEOVOL factor. By allowing for heterogeneous volatility factors, the GEOVOL model can be used in more general settings than geopolitical.

The increasing relevance of geopolitical risk and the need for quantifying it gave rise to an increasing interest among not only practitioners but also academics. Several researchers are now providing their own indicators of geopolitical risk. As for the methods used, textual analysis (Baker et al., 2016, Caldara and Iacoviello, 2018) is becoming very popular. Modelling multiplicative volatility factors using numerical methods is easy to implement and to replicate. Another appealing feature of the multiplicative decomposition proposed is that it implies a one-factor structure of the squared innovation covariance matrix to which factor or principal component analysis can be applied. This follows the

literature on one-factor models to explain return variances; See Trzcinka (1986), Connor and Korajczyk (1993) and Jones (2001) for references.

The modelling and estimation strategy of GEOVOL is applied to country equity indices from 1996 until 2019. The results indicate that GEOVOL spikes around the 9/11 attack, financial and economic crises, and political voting. The country indices are affected differently if the GEOVOL factor loadings differ across assets. Despite volatility shocks that affect all portfolios, some assets are more sensitive to the volatility shocks than others. Hence, there is a role for risk diversification.

Geopolitical risk is often the explanation for weak investment results. Conventional Markowitz style portfolios are predicted to have low volatility but they may be very sensitive to volatility shocks. Thus, assets that are not sensitive to volatility shocks are likely to be attractive in a portfolio because they diversify geopolitical risk. Consequently, the GEOVOL model also allows for a new criterion for portfolio optimality which complements the mean-variance efficiency by reducing the exposure to geopolitical risk.

The paper is organized as follows. In the following section, the baseline model of GEOVOL and the model with heterogeneous GEOVOL effects are presented and the estimation method is developed. Subsequently, in section 3, a new test of GEOVOL effects is proposed and its properties in finite samples are investigated using Monte Carlo simulations. An empirical application to country equity indices is provided in section 4 where the empirical results are discussed and compared to other existing measures. Section 5 develops implications for portfolio formation and section 6 explores the forecasting of geopolitical risk. Section 7 updates the estimates for the first 5 months of 2020 in order to view the COVID-19 period. Finally, section 8 concludes the paper.

2 The model of GEOVOL

It is well known that volatilities of financial returns tend to rise and fall together. Multiplicative volatility factors are introduced to explain the variation of all random variables in multivariate time series. They are assumed as the drivers of comovements of volatilities and, primarily, of shocks to volatilities. We start by representing the baseline model of GEOVOL and later the model is extended to accommodate heterogeneous impacts of GEOVOL on the individual volatilities.

The standard asset pricing model can be formulated for an $(N \times 1)$ vector of returns $\boldsymbol{r}_t \equiv (r_{1t}, \ldots, r_{Nt})'$ as,

$$\begin{aligned} \boldsymbol{f}_t &= \boldsymbol{w}_{t-1}' \boldsymbol{r}_t \\ \boldsymbol{r}_t &= r^f + \boldsymbol{\beta} \boldsymbol{f}_t + \mathsf{diag}\{\sqrt{\boldsymbol{h}_t}\}\boldsymbol{e}_t, \end{aligned} \tag{1}$$

where $\boldsymbol{\beta}$ is an $(N \times K)$ matrix of risk exposures, $\boldsymbol{f}_t$ is a $(K \times 1)$ vector of factors, $\boldsymbol{e}_t \equiv (e_{1t}, \ldots, e_{Nt})'$ is the vector of residuals from factors, and $\boldsymbol{h}_t \equiv (h_{1t}, \ldots, h_{Nt})'$ contains the conditional variances. We define $\mathsf{diag}\{\boldsymbol{a}\}$ as a matrix with the vector $\boldsymbol{a}$ on the

diagonal and 0 elsewhere. If model (1) is correctly specified and factors fully explain the cross sectional correlation, then $\boldsymbol{e}_t$ contains idiosyncratic returns and $\boldsymbol{h}_t$ idiosyncratic conditional variances. The standard assumptions on $\boldsymbol{e}_t$ state that the standardized residuals are uncorrelated in both time series and cross section with unit variances. Hence, if factors are sufficient to reduce the contemporaneous correlations to zero,

$$\mathbb{E}_{t-1}(\boldsymbol{e}_t \boldsymbol{e}_t') = \mathbb{I}. \tag{2}$$

Satisfying the assumptions does not imply that the elements of $\boldsymbol{e}_t$ are independent, only that they are uncorrelated. If they were independent, then all functions of the elements of $\boldsymbol{e}_t$ would also be independent and there would be no comovements of any kind. Consequently, the square (or absolute value) of $\boldsymbol{e}_t$ may be correlated in the cross section. Define ψ_{it} as a volatility shock in the univariate case as follows

$$\psi_{it} \equiv e_{it}^2 - 1 = \frac{(r_{it} - r^f - \boldsymbol{\beta}_i' \boldsymbol{f}_t)^2 - h_{it}}{h_{it}}, \tag{3}$$

where $\boldsymbol{\beta}_i$ is a column vector containing the elements of the ith row of matrix $\boldsymbol{\beta}$. The volatility shock ψ_{it} represents the proportional difference between the squared ith idiosyncrasy and its expectation. In univariate settings, the realized e_t^2, are on some dates bigger than one and on others smaller than one. If many assets have e_t^2 bigger than one at the same time, this can be interpreted as a common volatility shock which we associate with geopolitical news (because of its broad impact on many assets).

2.1 The GEOVOL statistical model

There is very strong evidence that the squared standardized residuals of returns net of factors are positively correlated. This observation in the time series was the motivation for the original ARCH model, see Engle (1982), and now the same observation in the cross section is the motivation for GEOVOL as a measure of geopolitical risk. GEOVOL will be high when squared standardized residuals are high for a wide range of assets. Thus it is a measure of the magnitude of shocks to volatility that are common to a collection of assets.

To estimate GEOVOL we must introduce parametric assumptions on the form of this relation. Let GEOVOL be represented by $\sqrt{x}$ where x is a $(T \times 1)$ vector of latent variables and let $\boldsymbol{s}$ be an $(N \times 1)$ vector of parameters interpreted as factor loadings satisfying the assumptions:

$$\mathbb{E}_{t-1}(x_t) = 1, \ \mathbb{E}_{t-1}(x_t - 1)^2 = v_t, \ x_t > 0, t = 1, \ldots, T, \tag{4}$$

$$\boldsymbol{\varepsilon}_t \sim IIN(0, 1), \tag{5}$$

$$s_i \in [0, 1], \ i = 1, \ldots, N. \tag{6}$$

We can specify a function $g(s_i, x_t)$ that is a data generating process for the random variables, e_{it} from x_t, s_i and ε_{it}. Throughout this paper we assume

$$e_{it} = \sqrt{g(s_i, x_t)}\varepsilon_{it}, \;\; g \equiv s_i x_t + 1 - s_i, \tag{7}$$

but other specifications are certainly possible. Specification (7) implies that g is non-negative with expected value 1 and therefore satisfies (2). It follows that

$$\begin{aligned}
\Psi_{ijt} &= \mathbb{E}_{t-1}\left[\varepsilon_{it}^2\varepsilon_{jt}^2\left(s_i s_j(x_t-1)^2 + (s_i+s_j)(x_t-1)+1\right) - 1\right] = s_i s_j v_t \\
\Psi_{iit} &= \mathbb{E}_{t-1}\left[\varepsilon_{it}^4\left(s_i^2(x_t-1)^2+1\right) - 1\right] = 3s_i^2 v_t + 2.
\end{aligned} \tag{8}$$

From (8) the sample covariance matrix can be constructed by averaging over t,

$$\begin{aligned}
\Psi &= \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} \boldsymbol{e}_t^2 \boldsymbol{e}_t^{2\prime}\right] = \frac{1}{T}\sum_{t=1}^{T}\Psi_t = \boldsymbol{s}\boldsymbol{s}'\bar{v} + D, \\
D &= \mathsf{diag}\{2s_i^2\bar{v} + 2\},
\end{aligned} \tag{9}$$

where $\bar{v} = (1/T)\sum_{t=1}^{T} v_t$. It is clear that Ψ is a factor matrix with x as the factor and $\boldsymbol{s}$ as the vector of factor loadings. Thus principle components analysis of the empirical version of Ψ will give preliminary estimates of both $\boldsymbol{s}$ and x.

2.2 Estimating the equal loading model

An important special case of the GEOVOL model is the version with equal factor loadings for all the assets. In this case we consider the vector $\boldsymbol{s} = \boldsymbol{\iota}$, a vector of ones. Equation (9) becomes

$$\Psi = \boldsymbol{\iota}\boldsymbol{\iota}'\bar{v} + 2(\bar{v}+1)\mathbb{I}. \tag{10}$$

The heteroskedasticity model of equation (7) becomes

$$e_{it} = \sqrt{g_t}\varepsilon_{it} = \sqrt{x_t}\varepsilon_{it}. \tag{11}$$

Maximum likelihood estimation of x_t will require simply a cross sectional estimate of observations on date t. Thus the likelihood function is

$$\begin{aligned}
L(x; \boldsymbol{e}) &= \sum_{t=1}^{T} L_t, \\
L_t &= -\frac{1}{2}\sum_{i=1}^{N}\left\{\log(g_t) + \frac{e_{it}^2}{g_t}\right\} = -\frac{1}{2}\left\{N\log(x_t) + \frac{\sum_{i=1}^{N} e_{it}^2}{x_t}\right\}.
\end{aligned} \tag{12}$$

The x which maximizes the log likelihood is just the maximum of each cross sectional estimate of x. Differentiating equation (12) with respect to x_t yields

$$\hat{x}_t = \frac{1}{N}\sum_{i=1}^{N} e_{it}^2. \tag{13}$$

This estimator does not depend upon v_t. It guarantees that x_t is positive and has a mean of one. It is presumably consistent as N goes to infinity with T finite or potentially increasing more slowly than N.

2.3 Estimating the GEOVOL model

The covariance matrix in (9) is observable and has information on the parameters of the model. However, it is not identified unless additional assumptions are made on the unknown parameters $\boldsymbol{s}$. Clearly if all the elements of $\boldsymbol{s}$ are multiplied by a scalar and v is divided by the square of this scalar, the covariances will be unchanged. This is hardly surprising as larger factor loadings would be associated with a factor with smaller variance.

Remark 2.1. *The covariances that identify the matrix of squared standardized innovations will be unchanged if each s_i is multiplied by a scalar γ and the variance of x is divided by γ^2.*

We therefore normalize the factor loadings by requiring that

$$\boldsymbol{s}'\boldsymbol{s} = 1. \tag{14}$$

This normalization is consistent with the requirement that the loadings are in the unit interval. It is also the normalization implicit in principal component analysis of (9).

To gain efficiency, we again want to use not just the unconditional covariances but also the observed heteroscedasticity relationships as in equation (7). These equations can be used to estimate $\boldsymbol{s}$ conditional on $\hat{x}$ by time series analysis or to estimate x conditional on $\hat{\boldsymbol{s}}$ from cross sectional analysis. The likelihood function can be written for this model as follows.

$$L(\boldsymbol{s}, x; \boldsymbol{e}) = -\frac{1}{2}\sum_{i=1,t=1}^{N,T}\left\{\log(g(s_i, x_t)) + \frac{e_{it}^2}{g(s_i, x_t)}\right\}. \tag{15}$$

This is not a classical likelihood function since x is considered a latent variable rather than a parameter. However, a common approach to estimation is to use the likelihood as if x were observed which is called *data augmentation* by Hastie et al. (2009). The

iteration solves the first order conditions

$$\begin{aligned} \frac{\partial L(\boldsymbol{s}, x; \boldsymbol{e})}{\partial s_i} &= 0, \\ \frac{\partial L(\boldsymbol{s}, x; \boldsymbol{e})}{\partial x_t} &= 0 \end{aligned} \tag{16}$$

sequentially until parameters are found that solve both jointly.

This algorithm can be interpreted as an Expectation-Maximization (EM) algorithm where the cross sectional regression estimates the unobserved value of x in the expectation step and then the time series regression maximizes the likelihood function conditional on the estimated latent variable. Since the expectation step is also a maximization, this is called a *Maximization-Maximization procedure* by Hastie et al. (2009). Each step therefore increases the likelihood function. The algorithm stops when the parameters become constant and hence the likelihood function has reached an extremum which can be verified to be a maximum. The initial estimates are given by principal components of the empirical Ψ matrix; See Trzcinka (1986), Connor and Korajczyk (1993) or Jones (2001) for one-factor models of return variances. For non-normal errors, we use the rank correlation matrix. With the initial loadings obtained from the first principal component, we are in the position to estimate the global volatility factor at each point in time using cross section analysis. Convergence occurs typically after something like 15 to 30 iterations. The normalizations are imposed after each step unless they are imposed in the step itself. As N gets large, the average correlation becomes smaller. Other choices for scaling can be derived from $\boldsymbol{s}'\boldsymbol{s} = aN$, where a is a constant between $1/N$ and 1. However, only when $a = 1/N$ we are guaranteed to have $s_i, i = 1, \ldots, N$, smaller than one and $g(s_i, x_t)$ positive for every (i, t).

Remark 2.2. *In each iteration,* $s_i, i = 1, \ldots, N$, *and* $x_t, t = 1, \ldots, T$, *are constrained to be, respectively, in the interval* $[0, 1]$ *and positive. Scaling is also imposed to guarantee* $\bar{x}_t = (1/T) \sum_{t=1}^{T} x_t = 1$ *and* $\boldsymbol{s}'\boldsymbol{s} = 1$.

2.4 Monte Carlo simulations

In the Monte Carlo simulations, the innovations $\boldsymbol{\varepsilon}_t$ are generated as random standard normal variables. The factor loadings $s_i, i = 1, \ldots, N$, are fixed, i.e., the same factor loadings are used across replications within the same dimension ($N = 10$ or $N = 50$). Table 1 shows the true values of the factor loadings $\boldsymbol{s}$ for $N = 10$ (upper panel) and $N = 50$ (lower panel). These values are generated by random draws from a uniform distribution in the interval $[0, 1]$.

The latent factor in the GEOVOL model can be assumed as either fixed or random. To generate data on replication r of GEOVOL, $x_{rt}, t = 1, \ldots, T$, is computed as the

Table 1: The true loadings $s_i, i = 1, \ldots, N$, used in the Monte Carlo simulations.

$N = 10$ ($\bar{s} = 0.253$)

0.013	0.092	0.370	0.097	0.488	0.072	0.068	0.394	0.469	0.471

$N = 50$ ($\bar{s} = 0.121$)

0.007	0.202	0.118	0.186	0.187	0.116	0.008	0.031	0.149	0.221
0.173	0.233	0.109	0.029	0.108	0.246	0.016	0.186	0.076	0.069
0.092	0.142	0.206	0.145	0.163	0.108	0.175	0.016	0.054	0.168
0.257	0.075	0.063	0.210	0.126	0.167	0.094	0.054	0.070	0.233
0.029	0.143	0.145	0.002	0.187	0.088	0.023	0.082	0.238	0.009

exponential of a random normal variable as follows:

$$x_{rt} = \exp(\phi_{rt}), \tag{17}$$

where ϕ_{rt} is drawn from a normal distribution with mean zero and variance v^2. Then, scaling is imposed to guarantee $1/T \sum_{t=1}^{T} x_{rt} = 1, r = 1, ..., R$. The number of replications, R, is 150.

To measure the accuracy of the estimator for any $x_t, t = 1, \ldots, T$, we compute an equivalent to the R^2 in regression analysis. This is done as follows:

$$\mathrm{MSE}_r(x - \hat{x}) = \frac{1}{T} \sum_{t=1}^{T} (x_{rt} - \hat{x}_{rt})^2 \tag{18}$$

$$\mathrm{MSE}_r(x - \bar{x}) = \frac{1}{T} \sum_{t=1}^{T} (x_{rt} - \bar{x})^2 \tag{19}$$

$$\mathsf{R}^2_x = \frac{1}{R} \sum_{r=1}^{R} \left\{ 1 - \frac{\mathrm{MSE}_r(x - \hat{x})}{\mathrm{MSE}_r(x - \bar{x})} \right\}, \tag{20}$$

where $\bar{x} = 1$. A similar procedure is applied to obtain R^2_s in which $\bar{s} = 1/N \sum_{t=1}^{N} s_i$. The higher the R^2 the more closely the estimated parameters match the simulated parameters.

The results for R^2_x and R^2_s using the baseline model and other statistics averaged across the replications are reported in Table 2. In the simulations, $\mathrm{v} = 2$. When assuming GEOVOL effects as random, the empirical variance of x_t, v, is averaged across replications. The first observation drawn from the simulations is that the average empirical correlation of squared innovations $\bar{\rho}_{\boldsymbol{e}^2}$ appears to decrease with the number of series which is consistent with the normalization (14). The results indicate that the estimator for the factor loadings and the factor can potentially be improved by increasing the number of, respectively, observations and series. In short samples, the precision of the estimator for $s_i, i = 1, \ldots, N$, can be adversely affected by an increasing number of series. Due to the normalization, the loadings become smaller as N increases and so more difficult to esti-

Table 2: Estimated R^2_s, R^2_x and other empirical statistics from the Monte Carlo simulations.

	Random x_t				Fixed x_t			
	$T = 1000$		$T = 5000$		$T = 1000$		$T = 5000$	
	$N = 10$	$N = 50$	$N = 10$	$N = 50$	$N = 10$	$N = 50$	$N = 10$	$N = 50$
R^2_s	0.924	0.640	0.981	0.926	0.924	0.632	0.983	0.924
R^2_x	0.810	0.885	0.820	0.895	0.785	0.901	0.787	0.888
$\bar{\rho}_{\boldsymbol{e}^2}$	0.147	0.076	0.154	0.096	0.175	0.109	0.165	0.101
v	28.79	27.61	37.53	36.66	38.13	38.13	35.82	35.82

mate accurately. Increasing the number of observations does not seem to have a significant impact on the accuracy of the estimator for $x_t, t = 1, \ldots, T$, since this also increases the number of parameters. Results are robust to assuming either random or fixed GEOVOL effects. Finally, the higher the average correlation of squared standardized innovations, the more accurate is the proposed estimator and so higher R^2_x and R^2_s are to be expected.

3 Testing for GEOVOL effects

An observable implication of the GEOVOL model is that, even though the elements of $\boldsymbol{e}_t$ are orthogonal in both times series and cross section, they may not be independent meaning that their squares can be correlated in the cross section. These comovements of $\boldsymbol{e}_t^2$ are induced by a common volatility factor, GEOVOL. It follows that detecting GEOVOL involves testing whether the squared standardized innovations are correlated.

Empirical evidence for GEOVOL is easy to find using the sample covariance matrix. The null hypothesis of no correlation in $\boldsymbol{e}_t$ is given by (2). Similarly, the null hypothesis for $\boldsymbol{e}_t^2$ is simply (9) with $\bar{v} = 0$, which implies

$$\mathbb{H}_0 : \Psi = 2\mathbb{I}. \tag{21}$$

The two in this equation is a result of assuming normality. Otherwise it would be the kurtosis of each return minus one. When the equal factor loading model is the alternative, all pairs of assets are affected by the same shock, and they will have the same correlation under the alternative. This is an equicorrelated panel and testing using the average correlation is sufficient for detecting GEOVOL. The "no GEOVOL effects" hypothesis holds when x_t is constant, that is when $v = 0$. In this setting, $\boldsymbol{e}_t$ are independent and no comovements of any kind can be observed across the standardized residuals. The null hypothesis $\mathbb{H}_0 : v = 0 \implies \bar{\rho}_{\boldsymbol{e}^2} = 0$, where $\bar{\rho}_{\boldsymbol{e}^2}$ denotes the average empirical correlation of $\boldsymbol{e}_t^2$. Under the alternative, the global volatility factor varies over time inducing comovements and positive correlations between the squared standardized residuals are observed. Hence, $\mathbb{H}_1 : v > 0 \implies \bar{\rho}_{\boldsymbol{e}^2} > 0$.

For $N(N-1)/2$ correlations, the test statistic becomes

$$\xi = \frac{\sqrt{\frac{NT}{(N-1)/2}} \sum_{i>j,j=1}^{T} \sum_{t=1}^{T} (e_{it}^2 - 1)(e_{jt}^2 - 1)}{\sum_{i=1}^{N} \sum_{t=1}^{T} (e_{it}^2 - 1)^2} \xrightarrow{d} N(0,1) \text{ under } \mathbb{H}_0. \tag{22}$$

We are in the position to state that there is very strong evidence that the squared standardized residuals of returns net of factors are positively correlated. The finite sample properties of this test as well as other similar tests for the average correlation in the literature (for comparison) are studied by means of Monte Carlo simulations in section 3.1. A brief description of these alternative tests is presented in Appendix A.

3.1 Finite sample properties

To investigate how the test for global volatility factors performs in finite samples, we run another Monte Carlo experiment with 100000 replications of samples with different dimensions ($N = 2, 5, 50$ and $T = 100, 1000$). For each replication, we compute the test statistic (22). Then, the empirical rejection frequencies are calculated for a particular nominal significance level ($\alpha = 0.01, 0.05$). For comparison, we also show the results for alternative test statistics described in the Appendix A. In the simulations, we assume $v = 0$ ($\bar{\rho}_{\boldsymbol{e}^2} = 0$) and $v > 0$ ($\bar{\rho}_{\boldsymbol{e}^2} > 0$) when studying, respectively, the size and the power of the test statistics.

Under the null hypothesis of no GEOVOL, $\boldsymbol{e}_t$ is simply generated by random standardized normal variables. These are naturally independent over time and across series which imply they are also uncorrelated. The results from the size simulations are summarized in Table 3. The size distortions are negligible even for the smallest samples ($N = 2$ or $N = 5$). In general, the empirical rejection frequencies tend to approximate the nominal significance levels when the number of either assets or observations increases.

For studying the power of the test, we start by considering the baseline model, i.e. the GEOVOL model with equal unit factor loadings or $s_i = 1$ for $i = 1, \ldots, N$. Then we generalize to the model with heterogeneous loadings on the global volatility factor. We generate x_t as in (17). The simulations results for power under the baseline GEOVOL model are shown in Table 4. Power is already high for low dimensional models ($N = 5$) regardless of the number of observations. It increases by adding assets to the sample and/or by considering longer times series.

The results from the power simulations with heterogeneous volatility factors are presented in Table 5. For each replication, the factor loadings $\boldsymbol{s}$ are drawn from a random uniform distribution in the interval $[0, 1]$ and normalized such that $\boldsymbol{s}'\boldsymbol{s} = 1$. Overall, power tends to decrease for the model with heterogeneous volatility factors when compared to the previous case with equal unit loadings. Moreover, there is almost no improvement in power by increasing the number of assets. The percentage of variance explained by GEOVOL is given by $\boldsymbol{s}'\boldsymbol{s}/N$. Due to the normalization $\boldsymbol{s}'\boldsymbol{s} = 1$, this ratio becomes $1/N$.

Table 3: Empirical rejection frequencies under $\mathbb{H}_0 : \bar{\rho}_{\boldsymbol{e}^2} = 0$.

DGP	T	N	t_{r_1}	t_{r_2}	t_{r_3}	t_z	ξ
				$\alpha = 0.01$			
1	100	2	0.022	0.022	0.022	0.022	0.019
2	100	5	0.016	0.018	0.017	0.019	0.017
3	100	50	0.014	0.018	0.015	0.019	0.016
4	1000	2	0.015	0.015	0.015	0.015	0.015
5	1000	5	0.013	0.013	0.013	0.013	0.013
6	1000	50	0.012	0.012	0.012	0.012	0.012
				$\alpha = 0.05$			
1	100	2	0.064	0.064	0.065	0.064	0.060
2	100	5	0.057	0.060	0.058	0.061	0.057
3	100	50	0.055	0.065	0.056	0.066	0.057
4	1000	2	0.056	0.056	0.056	0.056	0.055
5	1000	5	0.053	0.053	0.053	0.053	0.053
6	1000	50	0.052	0.052	0.052	0.052	0.052

Increasing N thus reduces the percentage of the variance of the standardized innovations explained by GEOVOL. This is then reflected in lower correlations of squared standardized innovations and consequently lower power. To increase power, an effective solution is to use long series (say at least $T = 1000$, about four years of daily observations).

4 An application to country equity indices

4.1 The data

The All Country World Index (ACWI) is a global equity index maintained by MSCI Inc. It is designed to measure the global equity-market performance, including stocks from developed and emerging markets. The index covers approximately 85% of the global investable equity opportunity set, including securities across mid- and large-cap size, style and sector segments. As of March 2019, the ACWI tracks 2,771 stocks with a total market capitalization of approximately 45,171 USD Billions. The top five constituent stocks in decreasing order of market capitalization are Apple, Microsoft Corporation, Amazon.com, Facebook and Johnson & Johnson. The top five countries by stock allocation are the United States (55.05% of the total index market capitalization), Japan (7.23%), the United Kingdom (5.15%), China (3.87%) and France (3.4%). This implies that all the other countries account for the remaining 25.3% of the total index market capitalization. Using the Global Industry Classification Standard method (GICS) taxonomy, most stocks are classified as Financials (16.59%), followed foremost by Information Technology (15.71%), Health care (11.65%) Consumer Discretionary (10.76%) and Industrials (10.44%).

The iShares MSCI World Exchange Traded Fund (ETF) tracks the ACWI and is maintained by Blackrock Inc. It seeks to track the investment results of an index which

Table 4: Empirical rejection frequencies under $\mathbb{H}_1 : \bar{\rho}_{e^2} > 0$ and GEOVOL model with heterogeneous effects.

DGP	T	N	t_{r_1}	t_{r_2}	t_{r_3}	t_z	ξ
			$\alpha = 0.01$				
1	100	2	0.219	0.219	0.220	0.219	0.183
2	100	5	0.365	0.382	0.368	0.387	0.389
3	100	50	0.651	0.679	0.654	0.685	0.705
4	1000	2	0.737	0.737	0.737	0.737	0.704
5	1000	5	0.978	0.979	0.978	0.979	0.986
6	1000	50	1.000	1.000	1.000	1.000	1.000
			$\alpha = 0.05$				
1	100	2	0.329	0.329	0.330	0.329	0.294
2	100	5	0.534	0.546	0.536	0.548	0.552
3	100	50	0.799	0.819	0.800	0.820	0.831
4	1000	2	0.820	0.820	0.820	0.820	0.794
5	1000	5	0.991	0.991	0.991	0.991	0.994
6	1000	50	1.000	1.000	1.000	1.000	1.000

v = 1 and $s_i, i = 1, \ldots, N$, drawn from $U(0, 1)$.

is composed by the shares of developed market companies around the world. The sample covers daily closing prices of 42 individual country ETFs traded in the New York Stock Exchange (NYSE). These are from March 18, 1996 to October 10, 2019 (total of 5933 observations). These assets have closing prices that are all observed at the same time so there is no asynchrony in returns. Because assets were introduced on different dates, this is an unbalanced panel. In fact, roughly 1/3 are missing at the beginning of the sample period. The currency is the U.S. dollar. The closing prices are converted into log-returns and, to avoid convergence problems in the estimation, extreme positive (negative) returns are truncated to +10% (10%). The summary statistics for each country equity ETF in the sample are reported in the Appendix B.

A factor model is estimated for each country equity ETF. A single factor, the cross section average of returns, is used and sufficient to capture all the common variation in the returns. The average correlation $\bar{\rho}_{\hat{e}} = -0.026$ which is not statistically different from zero. An AR(1) model is also considered when time dependence is observed in the first moment of the data. Given the strong evidence for the presence of ARCH effects, we consider a GARCH$(1, 1)$ model for modelling the second moment.

In fact, a first-order GARCH process is usually sufficient to capture the heteroskedastic behaviour of the time series in financial applications. The test statistics and $p-$values from Ljung-Box AR(1) and ARCH(1) tests can also be found in Appendix B.

For the sake of saving space, we are only presenting the averaged estimated residuals and volatilities across the 42 assets in the sample. These are shown in Figure 1. The daily cross-sectional mean residual is computed as $\bar{\hat{e}}_t = (1/N) \sum_{i=1}^{N} \hat{e_{it}}$ and the daily cross-sectional mean volatility as the square root of the mean variance $\sqrt{(1/N) \sum_{i=1}^{N} \hat{h_{it}}}$.

Table 5: Empirical rejection frequencies under $\mathbb{H}_1 : \bar{\rho}_{\boldsymbol{e}^2} > 0$ and the baseline GEOVOL model.

DGP	T	N	t_{r_1}	t_{r_2}	t_{r_3}	t_z	ξ
			$\alpha = 0.01$				
1	100	2	0.149	0.149	0.150	0.149	0.131
2	100	5	0.638	0.650	0.641	0.657	0.630
3	100	50	1.000	1.000	1.000	1.000	1.000
4	1000	2	0.684	0.684	0.684	0.684	0.678
5	1000	5	1.000	1.000	1.000	1.000	1.000
6	1000	50	1.000	1.000	1.000	1.000	1.000
			$\alpha = 0.05$				
1	100	2	0.264	0.264	0.266	0.264	0.245
2	100	5	0.807	0.813	0.809	0.815	0.797
3	100	50	1.000	1.000	1.000	1.000	1.000
4	1000	2	0.846	0.846	0.846	0.846	0.843
5	1000	5	1.000	1.000	1.000	1.000	1.000
6	1000	50	1.000	1.000	1.000	1.000	1.000

v = 0.5 and $s_i = 1, i = 1, \ldots, N$,.

The late 1990's and early 2000's, and the global financial crisis mark periods of very large shocks and very high volatility.

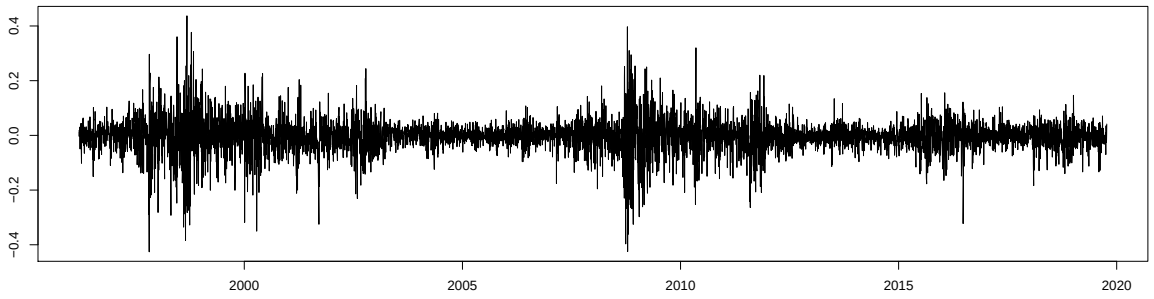
Then, the volatility standardized residuals are saved. The average correlation $\bar{\rho}_{\hat{\boldsymbol{e}}} = 0.061$. A test of whether this average correlation of the squared standardized residuals is significantly different from zero is given by $\xi = 111.8$ which should be a standard normal if $\bar{\rho}_{\boldsymbol{e}^2} = 0$. Even though the cross sectional correlation of standardized residuals is zero, there is strong statistical evidence that the squares are in fact positively correlated.

4.2 Estimated GEOVOL and GEOVOL loadings

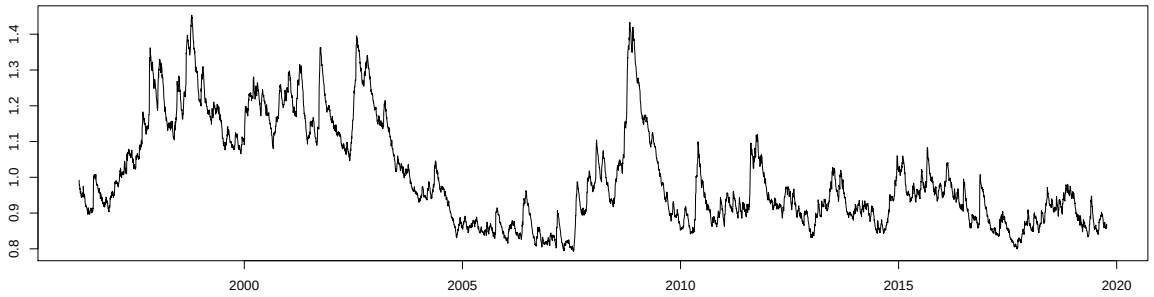
We now proceed to the estimation of the model with GEOVOL effects. Using the squared standardized residuals, we apply principal component analysis and record the loadings on the first principal component. The next steps include the estimation of $x_t, t = 1, \ldots, T$, using cross sectional heteroskedasticity regressions, and by imposing $x_t > 0$ for every t and re-scaling according to the formula

$$\bar{\hat{x}}_t \equiv \hat{x}_t / \bar{\hat{x}} \text{ where } \bar{\hat{x}} = \frac{1}{T} \sum_{t=1}^{T} \hat{x}_t. \tag{23}$$

The factor loadings $s_i, i = 1, \ldots, N$, are estimated using time series heteroskedasticity regressions, and by restricting $0 \leq s_i \leq 1$ for every i and the re-scaling $\hat{s}_i / \sqrt{\sum_{i=1}^{N} \hat{s}_i^2}$. Re-scaling means that, after each step, there is a normalization to guarantee that the mean of $x_t, t = 1, \ldots, T$, is 1 and $\boldsymbol{s}'\boldsymbol{s} = 1$. These two steps complete the first iteration. To gain efficiency, we iterate the two steps until convergence. For the sample analyzed,



(a) Daily mean residuals



(b) Daily mean volatilities

Figure 1: The estimated daily mean residuals from a factor model (upper panel) and daily mean volatilities (bottom panel) averaged across the 42 country equity ETFs from March 18, 1996 to October 10, 2019.

the optimization algorithm converged after 11 iterations.

At this point we have estimates of both GEOVOL and the GEOVOL factor loadings. After sorting GEOVOL, its largest estimated values are shown in Table 6. For comparison, we also report the values of the average returns across the country assets on the same day. It is clear the role of financial and political events in GEOVOL. These show that the highest levels of GEOVOL over the last two decades happened on the day after the United Kingdom European Union membership referendum with its decision in favor of the *Brexit* on June 23, 2016, the two days after the United States presidential election on November 8, 2016, the day which the NYSE opened after the 9/11 terrorist attacks, the collapse of the Chinese stock market before the financial crisis and again in 2015, the day after the first round of the French presidential on April 23, 2017, the global stock market crash in October, 1997 (after the Asian financial crisis) with massive losses but sharp rebounds (completely unexpected on October 28, 1997), the stock market downturn in 2002, the day after oil prices crashed on November 27, 2014. Clearly some of these are political and some are economic events, but they all move the market. From this table, we can also see that the factor return on each day with high GEOVOL. Most of these are negative suggesting that big volatility is associated with negative news. However, there

Table 6: Largest values of the estimated GEOVOL factor ($\hat{x}_t$) and corresponding dates. The average return over the cross section ($\bar{r}_t$) and the return on the S&P 500 index (r_t^{spx}) on the same day are also shown.

	$\hat{x}_t$	$\bar{r}_t$	r_t^{spx}
2016-06-24	29.764	−7.002	−3.658
2016-11-09	29.121	−1.229	1.102
2001-09-17	26.799	−6.722	−5.047
2007-02-27	25.460	−6.002	−3.534
2017-04-24	21.699	2.230	1.078
1997-10-28	20.870	3.910	4.989
1997-04-21	19.971	0.878	−0.782
2015-08-24	19.294	−4.098	−4.021
2011-08-05	17.506	0.399	−0.058
2008-10-10	16.395	−2.648	−1.183
1999-01-04	15.893	2.784	−0.092
2016-11-10	15.444	−1.645	0.195
2014-11-28	15.250	−1.623	−0.255
2010-05-10	15.042	6.495	4.303
2018-04-09	14.493	0.379	0.333

are some positive events as well.

These results allow an assessment of the level of geopolitical risk over time. From plotting GEOVOL we can see how many events occurred and how big the shocks were. Monthly averages of GEOVOL are presented in Figure 2. The main events are labeled. It is clear the dimension of 9/11 terrorist attack to the World Trade Center. After that comes the great financial crisis and the sovereign debt crises. Interestingly, the level of geopolitical events in the recent period is rather low. This is contrary to common wisdom consistent with other measures which considers the current period to be extremely perilous. These additional measures are relatively high now in contrast to what is predicted by GEOVOL. However, it is widely recognized that we are currently in a low volatility period in financial markets.

The measure used in this paper is based on the common volatility shocks to financial markets. Other measures are usually based either on textual analysis or expert opinion. Thus GEOVOL is more likely to reflect what happened and the other measures what people are worried might happen.

The global economic policy uncertainty index (GPU) of Baker et al. (2016) is a newspaper- based index constructed using key terms pertaining to uncertainty, the economy, and policy. It is a monthly index developed for analyzing uncertainty in the United States (US) and extended to other countries and to different policy dimensions. The text search method applied to the US index is the same to each of the other eleven individual countries. Their results indicate higher economic policy uncertainty near presidential elections, wars, terrorist attacks and major fiscal debates. As these country indices are constructed based on textual analysis of newspapers distributed in a particular country,

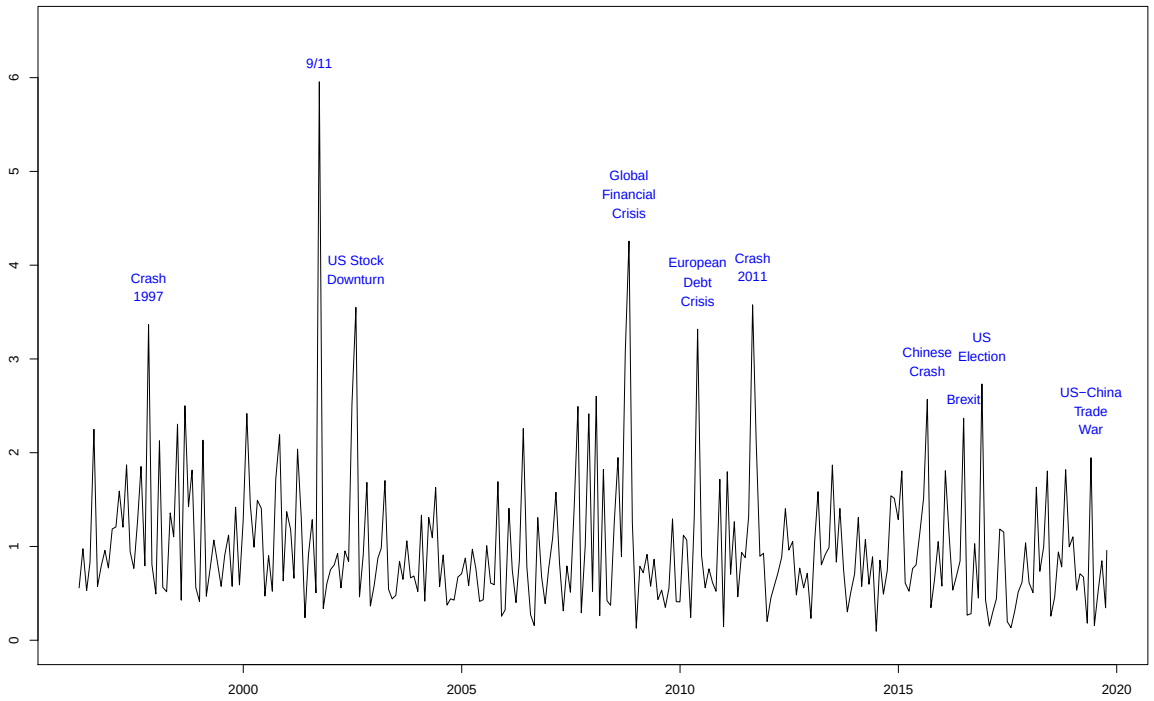


Figure 2: The estimated GEOVOL (monthly averages).

they tend to capture uncertainty driven by country-specific factors rather than common global factors. The geopolitical risk index (GPR) of Caldara and Iacoviello (2018) is the monthly ratio of the number of articles related to geopolitical tensions to the total number of articles in eleven newspapers published in the US, the United Kingdom and Canada since 1985. By using textual-analysis, the index may be biased by the category of words selected. In the GPR, for instance, the words picked are associated with explicit mentions of geopolitical risk and military-related tensions as well as nuclear tensions, war threats and terrorist threats. It is presumably a better indicator of military risk. The correlation between GEOVOL and the first differences of the GPU and the GPR is, respectively, 0.401 and 0.156.

Finally, the estimated factor loadings for the countries analyzed are reported in Table 7 in descending order of magnitude. The factor used in the mean equations (the cross sectional average of returns) is also included in the sample. This is natural because it is a weighted average of individual countries which are each exposed to GEOVOL and presumably the factor would be especially exposed. Furthermore, the factor is by assumption uncorrelated with the other idiosyncratic shocks. Results indicate that the countries of France, Netherlands or Germany have higher loadings on the global volatility factor and so are more sensitive to geopolitical shocks. On the other hand, countries with lower factor loadings such as Poland, New Zealand or Pakistan appear to be less sensitive. The factor return has the second highest loading on the GEOVOL factor. To diversify geopolitical risk and reduce the impact of GEOVOL on portfolio variance, higher (lower)

Table 7: Estimated GEOVOL factor loading $\hat{s}_i$ for each country ETF.

Average	0.239	Portugal	0.154	Greece	0.128
France	0.239	Mexico	0.151	Ireland	0.126
Netherlands	0.221	Hong Kong	0.145	Australia	0.125
Germany	0.216	China	0.140	Turkey	0.125
Spain	0.215	United Kingdom	0.140	Philippines	0.124
Italy	0.211	Japan	0.140	Norway	0.114
Belgium	0.202	Denmark	0.139	Israel	0.110
Malaysia	0.190	Taiwan	0.138	Russia	0.103
Thailand	0.186	United States	0.137	India	0.094
Austria	0.175	Colombia	0.136	Peru	0.083
Korea	0.167	Chile	0.135	Poland	0.080
Indonesia	0.163	Brazil	0.134	New Zealand	0.062
Singapore	0.160	Canada	0.132	Pakistan	0.014
Sweden	0.157	Switzerland	0.130		
Finland	0.154	South Africa	0.129		

weights should be given to assets with smaller (larger) loadings on the global volatility factor, i.e., on GEOVOL. This is discussed in section 5.

As a measure of the goodness of fit, we standardize the residuals by using not only the conditional variances but also GEOVOL. If this common global factor captures the common global shocks in the standardized residuals, then the standardized residuals should be independent and no comovements should be observed. In practical terms, this means that $\boldsymbol{\varepsilon}_t^2$ should be uncorrelated. The test for detecting GEOVOL is thus applied to the square of the double standardized residuals (notice that $\boldsymbol{\varepsilon}_t^2 = \boldsymbol{e}_t^2/\boldsymbol{g}_t$, where $\boldsymbol{g}_t \equiv (g(s_1, x_t), \ldots, g(s_N, x_t))'$) using not only the conditional variances but also GEOVOL. Thus, the null hypothesis being tested becomes $\mathbb{H}_0' : \bar{\rho}_{\boldsymbol{e}^2/\boldsymbol{g}_t} = 0$. The empirical average correlation $\bar{\rho}_{\hat{\boldsymbol{e}}^2/\hat{\boldsymbol{g}}_t} = -0.002$, where $\hat{\boldsymbol{g}}_t \equiv (g(\hat{s}_1, \hat{x}_t), \ldots, g(\hat{s}_N, \hat{x}_t))'$ and for which the test statistic $\xi = -2.711$. This failure to reject the null hypothesis supports GEOVOL as a measure of the magnitude of common global shocks as it effectively captures the contemporaneous correlation in the squared standardized residuals.

Even though our empirical results are useful for tracking geopolitical risk as just shown, our approach does not exclude applications to assets of different classes or sectors within a nation. Another empirical application to nine US sector equity ETFs shows remarkably similar results for the largest GEOVOL events and for its monthly averages. These results are available upon request.

5 Portfolio implications

Hedging geopolitical risk has practical benefits to investors and firms. As investors look beyond their home market for the global investment opportunity set, it matters from where risk is coming and to where it spreads. However, risk arising from a geopolitical event is potentially very dangerous to investors since a single event can result in increased

volatility for all assets in a portfolio or all lines of business for a conglomerate firm. Conventional diversification does not reduce its impact. However, if the GEOVOL factor loadings on assets differ, it is possible to reduce (but not eliminate) the exposure to geopolitical risk. Thus a new criterion for portfolio optimality is introduced which we label *risk diversification*.

As a simple illustration consider two uncorrelated assets with the same variance and expected return but one is exposed to geopolitical risk. The Markowitz portfolio is equally weighted. Using the following specification, we see that the kurtosis of the portfolio depends on the variance of GEOVOL.

$$r_{1t} = \sqrt{x_t}\varepsilon_{1t}, \quad r_{2t} = \varepsilon_{2t}, \quad \varepsilon_{1t}, \varepsilon_{2t} \sim IN(0, \sigma^2), \quad x_t \sim D(1, \kappa) \tag{24}$$

$$V(\pi_t^{Markowitz}) = \frac{\sigma^2}{2}, \quad \frac{\mathbb{E}\left[(\pi_t^{Markowitz})^4\right]}{V^2(\pi_t^{Markowitz})} = 3 + \frac{3}{4}\kappa \tag{25}$$

Clearly, by reducing the exposure to the first asset, the kurtosis of the portfolio can be reduced to zero but the variance will be increased. If the variance of x is large, indicating a long right tail in this non-negative random variable, it is likely to be preferable to reduce exposure to the first asset.

The variance-covariance matrix of the vector of returns can be computed from (1) conditional on x_t. Because factors are linear combinations of asset returns, they will also be affected by x_t. We start by rewriting the equations for an $(N \times 1)$ vector of returns r_t and a $(K \times 1)$ vector of factors f_t as a function of x_t:

$$r_{jt} = \sum_{k=1}^{K} \beta_{jk} f_{kt} + \sqrt{h_{jt}(s_j x_t + 1 - s_j)}\varepsilon_{jt} \tag{26}$$

$$f_{kt} = \mu_k + \sqrt{h_{kt}(s_k x_t + 1 - s_k)}\varepsilon_{kt}. \tag{27}$$

The variance of factor k conditional on the past and on x is given by

$$V_{t-1}(f_{kt}|x_t) = h_{kt} s_k x_t + h_{kt}(1 - s_k) \tag{28}$$

which depends linearly on x. On average, the effect of x on the variance of factor k is:

$$V(f_{kt}|x) = \sigma_k^2 s_k (x - 1) + \sigma_k^2. \tag{29}$$

denoting the unconditional variance of asset k as σ_k^2. Thus the average exposure to geopolitical risk from investing in factor k is $\sigma_k^2 s_k$. We define for factor k:

$$\text{Variance weighted GEOVOL loading: } \tilde{s}_k = s_k \sigma_k^2. \tag{30}$$

A similar measure can be constructed for individual stocks. Because the factor error and the idiosyncratic error are uncorrelated, the variance is obtained by squaring and

summing these terms as follows:

$$V_{t-1}(r_{jt}|x) = h_{jt}(s_j(x_t - 1) + 1) + \sum_{k=1}^{K} \beta_{jk}^2 h_{kt}(s_k(x_t - 1) + 1) \tag{31}$$

and letting $\tilde{s}_j = s_j \sigma_j^2$,

$$V(r_{jt}|x) = (x - 1) \left[\tilde{s}_j + \sum_{k=1}^{K} \beta_{jk}^2 \tilde{s}_k \right] + \left[\sigma_j^2 + \sum_{k=1}^{K} \beta_{jk}^2 \sigma_k^2 \right]. \tag{32}$$

The covariance matrix of returns can be expressed in matrix notation. For K orthogonal factors and N individual assets the notation can be expressed as follows:

$$\begin{aligned}
\boldsymbol{\beta} &: (N \times K) \text{ factor loadings,} \\
\tilde{\boldsymbol{s}}_f &: (K \times 1) \text{ variance weighted GEOVOL loadings of factors,} \\
\boldsymbol{\sigma}_f^2 &: (K \times 1) \text{ unconditional variance of factors,} \\
\tilde{\boldsymbol{s}}_i &: (N \times 1) \text{ variance weighted GEOVOL loadings of idiosyncracies,} \\
\boldsymbol{\sigma}_i^2 &: (N \times 1) \text{ unconditional variance of idiosyncracies}
\end{aligned}$$

and

$$V(r|x) = (x - 1) \left[\boldsymbol{\beta} \mathsf{diag}\{\tilde{\boldsymbol{s}}_f\} \boldsymbol{\beta}' + \mathsf{diag}\{\tilde{\boldsymbol{s}}_i\} \right] + \boldsymbol{\beta} \mathsf{diag}\{\boldsymbol{\sigma}_f^2\} \boldsymbol{\beta}' + \mathsf{diag}\{\boldsymbol{\sigma}_i^2\}. \tag{33}$$

Notice that this is just the unconditional variance of a factor model when $x = 1$. The expression is linear in x so the impact of x on portfolio variance is easy to define.

Portfolios with weights $\boldsymbol{w}$ can be constructed based on this covariance matrix and an assumed vector of expected returns, $\boldsymbol{\mu}$. A portfolio with weights $\boldsymbol{w}$ will thus have variance conditional on x and given by

$$V(\pi|x) = \boldsymbol{w}' \left[(x - 1) \left(\boldsymbol{\beta} \mathsf{diag}\{\tilde{\boldsymbol{s}}_f\} \boldsymbol{\beta}' + \mathsf{diag}\{\tilde{\boldsymbol{s}}_i\} \right) + \boldsymbol{\beta} \mathsf{diag}\{\boldsymbol{\sigma}_f^2\} \boldsymbol{\beta}' + \mathsf{diag}\{\boldsymbol{\sigma}_i^2\} \right] \boldsymbol{w}. \tag{34}$$

By considering both a variance criterion and a geopolitical risk criterion, a more stable portfolio can be constructed. The optimization could be as follows:

$$\begin{aligned}
&\text{Max } \boldsymbol{w}' \boldsymbol{\mu} \\
&\text{subject to } \boldsymbol{w}' \left[\boldsymbol{\beta} \mathsf{diag}\{\boldsymbol{\sigma}_f^2\} \boldsymbol{\beta}' + \mathsf{diag}\{\boldsymbol{\sigma}_i^2\} \right] \boldsymbol{w} < \theta_1 \\
&\text{and } \boldsymbol{w}' \left[\boldsymbol{\beta} \mathsf{diag}\{\tilde{\boldsymbol{s}}_f\} \boldsymbol{\beta}' + \mathsf{diag}\{\tilde{\boldsymbol{s}}_i\} \right] \boldsymbol{w} < \theta_2.
\end{aligned} \tag{35}$$

for parameters (θ_1, θ_2). If the second constraint is binding, then the portfolio will reduce the impact of geopolitical shocks by over-weighting assets and factors with smaller $\tilde{s}$.

6 Forecasting GEOVOL

The structure of this model suggests that GEOVOL is not predictable. However, there may be some channels for prediction. Although the mean of GEOVOL is always 1, the variance of GEOVOL can change over time. In this case, the probability distribution changes its tail properties over time. The higher the variance the greater the probability of a large geopolitical event. The probability can be measured by standard tail constructs such as the VaR which might be called $GVaR$ for Geopolitical Value at Risk. Because x is a variance, it is defined for probability α as

$$P(x > GVaR^2) = \alpha. \tag{36}$$

The tail properties of non-negative random variables with a mean of one are naturally investigated by assuming a distribution and here we use the gamma. There are only two parameters to adjust - the mean and variance of the random variable in the forecast period. Then quantiles can be computed; we will use the 1% quantile for $GVaR$.

In order to calculate the $GVaR$ it is natural to assume that v_t has a rather smooth time series which can be estimated by computing the conditional variance of x using some type of volatility model. The process does not look very much like a typical GARCH model because volatility does not cluster. Thus a GARCH model will indicate little persistence and give predictions which are sensitive to the previous event.

There is however an additional effect which is observed in empirical models which suggests an extension of the statistical specification. There is autocorrelation in the estimated series of x which suggests that

$$\mathbb{E}(x_t|x_{t-1}) \neq 1. \tag{37}$$

The autocorrelation in the estimated x would also be available for predicting geopolitical risk.

From equation (4) it is assumed that

$$\mathbb{E}_{t-1}(x_t) = 1. \tag{38}$$

However, this does not contradict equation (37) since lagged x is not in the public information set. More precisely, x_t is not measurable with respect to r_t. A slightly weaker version of equation (38) can extend this point. A natural assumption is:

$$\mathbb{E}^j_{t-1}(x_t) = 1 \text{ for } j = 1, \ldots, N, \text{ where } \mathbb{E}^j_{t-1}(x_t) \equiv \mathbb{E}(x_t|r_{j,t-1}, \ldots, r_{j,1}). \tag{39}$$

The expectation of x is one from each of the N data sets, one at a time. This still allows the possibility of forecasting x based on joint evidence from the collection of all assets.

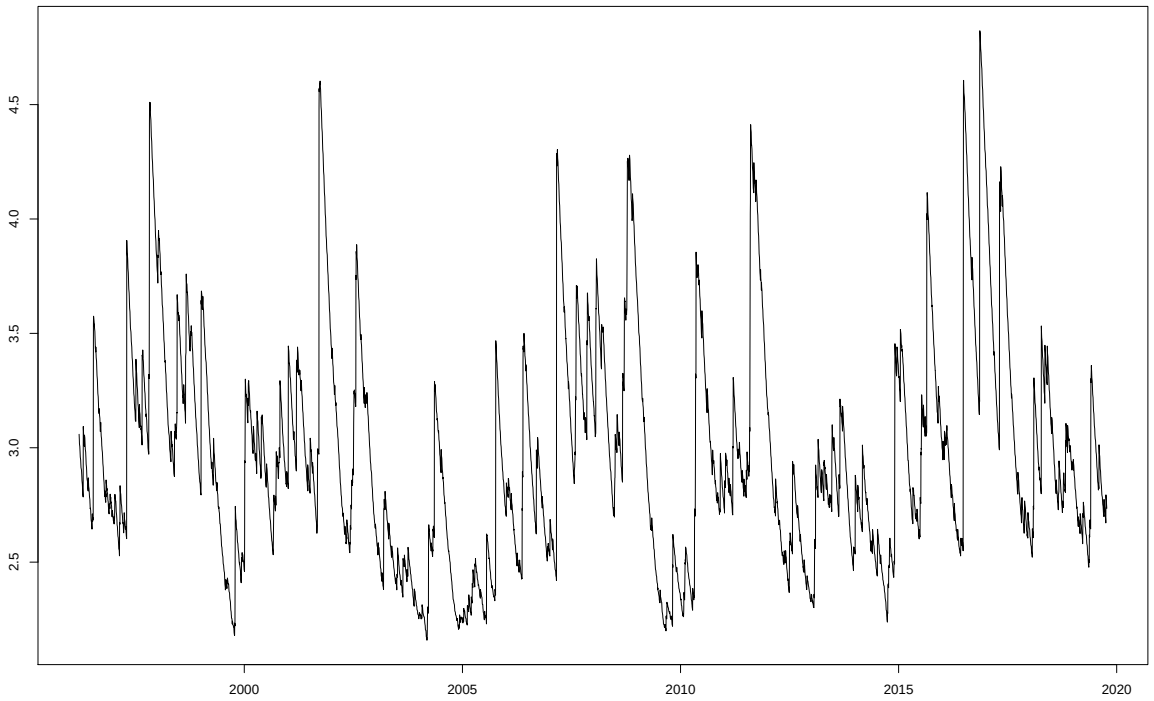


Figure 3: $GVaR^2$: Geopolitical Value at Risk for GEOVOL at 1% level.

Nevertheless from the point of view of each of these N information sets,

$$\mathbb{E}_{t-1}^{j}(\boldsymbol{e}_t\boldsymbol{e}_t') = \mathbb{I} \text{ for all } j = 1, \ldots, N. \tag{40}$$

Thus there is the possibility of forecasting x based on a multivariate indicator such as x_{t-1} even though no single asset or factor could provide this information.

In the data set used in this paper, the time series first order autocorrelation of $\hat{x}_t$ is 0.222 and it declines monotonically to 0.060 after 5 days. We estimate the mean of $(\hat{x}_t - 1)$ as an ARMA(1,1) with a variance exponentially smoothed with parameter $\lambda = 0.995$ implying a half life of six months.

The 1% upper tail quantile of geopolitical volatility one day in the future assuming that it is a gamma random variable with this mean and variance, is called $GVaR2$. The square root is plotted in Figure 3. It should be interpreted as a proportional increase in all asset volatilities that will only be exceeded one day in 100.

It is clear that asset volatilities more than 3 times current levels are on average likely to be exceeded 1% of the time. At the highest levels, this is 10 times. The risk of high geopolitical events fluctuates over time. It was high during 9/11, the financial crisis and in 2016 but falls through October 2019 which is the end of this data set and which was a time of low financial volatility.

7 Update 2020

This model has been updated through the first 5 months of 2020 - the first part of the COVID-19 pandemic. The overall estimates are very similar. However there are several new dates that are in the top 20 events of the last three decades. These dates are March 9, 12, 13 in 2020. The pandemic has sparked dramatic increases in financial market volatility and these three events are among the most severe and universal shocks to the financial system. Interestingly, March 9 is also the day on which Russia and Saudia Arabia failed to limit oil production and oil futures prices fell dramatically.

This model is now publicly available on *V-Lab*. It is now possible to see the current state of GEOVOL and the historical estimates from prior data sets. Further research is underway to examine the same model for exchange rates, sovereign bond markets and broad classes of asset prices.

8 Conclusions

When factors capture all the cross sectional correlation of financial returns, standardized innovations are orthogonal. Despite being uncorrelated, they may not be independent. The key motivation for the GEOVOL model comes from the fact that square standardized innovations are correlated in the cross section. A novel explanation for why idiosyncratic volatilities comove is thus provided and a new way to formulate multiplicative factors for volatility is introduced.

The model assumes a multiplicative decomposition of the innovations to idiosyncratic volatilities. The global volatility factor, or GEOVOL, captures the common variation in the innovations to volatilities. It is a measure of the magnitude of common volatility shocks and, due to its broad impact on a wide range of assets, is linked to geopolitical news. In order to account for heterogeneous GEOVOL effects, assets are allowed to have different loadings on the GEOVOL factor.

Using country equity indices, we identify GEOVOL effects arising from political, financial, economic and terrorist events. Country indices around the world have different loadings on the global volatility factor meaning they have different exposures to GEOVOL. Risk diversification can thus be improved by including GEOVOL as an additional criterion in portfolio optimization.

Acknowledgements

Part of this work was developed while the second author was visiting the New York University Stern School of Business. The financial support provided by the Luso-American Development Foundation and the Portuguese Foundation for Science and Technology (SFRH/BD/109539/2015) is gratefully acknowledged. This research was supported by the Volatility and Risk Institute at NYU Stern School of Business and the authors are indebted

to the Global Risk Institute, the Sloan Foundation, the National Science Foundation, the Norwegian Finance Institute and many individual donors.

A Alternative test statistics

For averaging correlation coefficients, we propose three alternative methods based on $z-$transformations. The idea is to transform each sample correlation coefficient $r_i, i = 1, \ldots, m$, where $m = N(N-1)/2$ is the number of unique correlation coefficients, to a Fisher's z and then take the average. The average $z-$transformed correlation has standard deviation equal to $1/\sqrt{m(T-3)}$ and is computed as follows:

$$\bar{z} = \frac{1}{m}\sum_{i=1}^{m} z_i, \tag{41}$$

where $z_i = 0.5\ln\{(1+r_i)/(1-r_i)\}$ follows a normal distribution with standard deviation $1/\sqrt{T-3}$. The test statistic to check $\mathbb{H}_0 : \bar{z} = 0$ is simply given by

$$t_z = \frac{\bar{z}}{1/\sqrt{m(T-3)}} \tag{42}$$

which follows a standard normal distribution under the null hypothesis.

In the simplest test, we use the arithmetic average of the correlation coefficients as follows:

$$\bar{r}_1 = \frac{1}{m}\sum_{i=1}^{m} r_i. \tag{43}$$

Because this estimator for the each of the alternative average correlation is negatively biased, we can back-transform $\bar{z}$ and test using the average correlation $\bar{r}_2$ which equals

$$\bar{r}_2 = \frac{e^{2\bar{z}}-1}{e^{2\bar{z}}+1}. \tag{44}$$

This is a less biased (but still with positive bias) alternative but with larger standard deviation. Thus, a superior estimator can be obtained by

$$\bar{r}_3 = \frac{1}{m}\sum_{i=1}^{m}\left\{r_i + \frac{r_i(1-r_i^2)}{2(T-3)}\right\}. \tag{45}$$

To test the null hypothesis that each of the alternative average correlations is zero, we compute the statistic

$$t_{r_i} = \frac{\bar{r}_i\sqrt{m(T-2)}}{\sqrt{1-\bar{r}_i^2}}, \tag{46}$$

which has a t distribution with $T-2$ degrees of freedom under the null hypothesis, $\mathbb{H}_0 : \bar{r}_i = 0, i = 1, 2, 3$.

B Summary statistics

Table 8: Summary statistics for the country index returns. Results from the Ljung-Box (L.B.) AR(1) and ARCH(1) tests are also shown.

	Austria	Australia	Belgium	Brazil	Canada	Switzerland
Min.	−10	−10	−10	−10	−10	−8.627
Mean	0.019	0.028	0.022	0.032	0.029	0.026
Max.	10	10	10	10	10	10
S.D.	1.623	1.631	1.512	2.315	1.405	1.354
Rob. Kr.	0.323	0.220	0.269	0.165	0.325	0.202
Rob. Sk.	0.099	0.094	0.074	−0.018	0.041	0.074
L.B. AR(1)	73.825	197.743	108.594	1.612	123.447	219.805
$p-$value	0.000	0.000	0.000	0.204	0.000	0.000
L.B. ARCH(1)	462.447	348.429	701.590	158.336	214.923	276.313
$p-$value	0.000	0.000	0.000	0.000	0.000	0.000
	Chile	China	Colombia	Germany	Denmark	Spain
Min.	−10	−8.568	−8.591	−10	−7.835	−10
Mean	−0.002	0.012	−0.032	0.019	0.048	0.027
Max.	10	8.955	8.583	10	4.351	10
S.D.	1.535	1.476	1.474	1.609	1.039	1.693
Rob. Kr.	0.110	0.089	0.392	0.206	0.116	0.212
Rob. Sk.	−0.021	0.015	−0.048	0.007	−0.002	0.030
L.B. AR(1)	2.025	6.534	5.284	75.464	11.328	35.870
$p-$value	0.155	0.011	0.022	0.000	0.001	0.000
L.B. ARCH(1)	100.262	225.558	45.025	314.171	15.345	343.532
$p-$value	0.000	0.000	0.000	0.000	0.000	0.000
	Finland	France	United Kingdom	Greece	Hong Kong	Indonesia
Min.	−10	−10	−10	−10	−10	−10
Mean	0.030	0.024	0.018	−0.008	0.018	0.011
Max.	6.654	10	10	10	10	9.265
S.D.	1.208	1.566	1.418	2.491	1.760	1.700
Rob. Kr.	0.161	0.212	0.204	0.222	0.243	0.111
Rob. Sk.	0.041	0.036	0.064	−0.005	0.045	−0.023
L.B. AR(1)	18.270	117.353	202.508	0.000	67.455	10.105
$p-$value	0.000	0.000	0.000	0.992	0.000	0.001
L.B. ARCH(1)	35.919	164.996	330.214	148.655	584.206	127.678
$p-$value	0.000	0.000	0.000	0.000	0.000	0.000

Table 8: Continued from previous page.

	Ireland	Israel	India	Italy	Japan	Korea
Min.	−10	−10	−6.872	−10	−10	−10
Mean	0.035	0.010	0.014	0.016	0.002	0.023
Max.	8.799	10	6.265	10	10	10
S.D.	1.318	1.375	1.352	1.716	1.460	2.034
Rob. Kr.	0.179	0.225	0.073	0.191	0.112	0.165
Rob. Sk.	0.024	0.022	0.046	0.003	0.009	−0.022
L.B. AR(1)	77.061	11.687	0.451	31.882	21.529	19.798
$p-$value	0.000	0.001	0.502	0.000	0.000	0.000
L.B. ARCH(1)	84.907	71.961	109.509	113.681	176.146	193.184
$p-$value	0.000	0.000	0.000	0.000	0.000	0.000
	Mexico	Malaysia	Netherlands	Norway	New Zealand	Peru
Min.	−10	−10	−10	−9.774	−8.107	−10
Mean	0.030	0.001	0.022	0.007	0.045	0.023
Max.	10	10	10	5.863	6.314	9.792
S.D.	1.855	1.801	1.528	1.333	1.079	1.310
Rob. Kr.	0.138	0.304	0.254	0.172	−0.018	0.139
Rob. Sk.	0.017	0.012	0.012	0.020	0.023	−0.030
L.B. AR(1)	7.371	9.108	160.988	55.836	7.534	14.702
$p-$value	0.007	0.003	0.000	0.000	0.006	0.000
L.B. ARCH(1)	211.365	536.415	243.279	157.314	4.608	96.069
$p-$value	0.000	0.000	0.000	0.000	0.032	0.000
	Philippines	Pakistan	Poland	Portugal	Russia	Sweden
Min.	−8.388	−5.506	−10	−10	−10	−10
Mean	0.016	−0.055	0.004	−0.012	0.005	0.028
Max.	6.950	6.567	8.630	5.449	10	10
S.D.	1.318	1.268	1.693	1.278	1.929	1.908
Rob. Kr.	0.105	0.237	0.117	0.191	0.122	0.259
Rob. Sk.	−0.024	−0.082	−0.058	−0.029	−0.017	0.079
L.B. AR(1)	5.695	0.977	0.076	5.718	27.865	95.245
$p-$value	0.017	0.323	0.783	0.017	0.000	0.000
L.B. ARCH(1)	37.654	7.045	43.877	20.294	196.910	390.750
$p-$value	0.000	0.008	0.000	0.000	0.000	0.000
	Singapore	Taiwan	Thailand	Turkey	United States	South Africa
Min.	−10	−10	−10	−10	−10	−10
Mean	0.009	0.007	0.030	−0.012	0.032	0.033
Max.	10	10	10	10	10	10
S.D.	1.777	1.852	1.709	2.375	1.192	2.050
Rob. Kr.	0.321	0.248	0.210	0.163	0.329	0.111
Rob. Sk.	0.016	−0.013	−0.023	0.009	−0.004	−0.009
L.B. AR(1)	85.127	29.068	0.213	0.046	36.018	7.381
$p-$value	0.000	0.000	0.645	0.830	0.000	0.007
L.B. ARCH(1)	462.481	281.786	49.424	168.386	231.794	80.013
$p-$value	0.000	0.000	0.000	0.000	0.000	0.000

References

Ang, A., R.J. Hodrick, Y. Xing, and X. Zhang, "The cross-section of volatility and expected returns," *The Journal of Finance*, 2006, *61*, 259–299.

Baker, S.R., N. Bloom, and S.J. Davis, "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 2016, *131*, 1593–1636.

Caldara, D. and M. Iacoviello, "Measuring Geopolitical Risk," 2018. Board of Governors of the Federal Reserve System: International Finance Discussion Paper 1222.

Connor, G. and R.A. Korajczyk, "A test for the number of factors in an approximate factor model," *Journal of Finance*, 1993, *48*, 1263 –1291.

_ , R.A. Korajczyk, and O. Linton, "The common and specific components of dynamic volatility," *Journal of Econometrics*, 2006, *132*, 231 – 255.

Engle, R.F., "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 1982, *50*, 987—1007.

Hastie, T., R. Tibshirani, and J.Friedman, "*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*," 2009. 2nd edition. New York City, USA: Springer.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh, "The common factor in idiosyncratic volatility: Quantitative asset pricing implications," *Journal of Financial Economics*, 2016, *119*, 249–283.

Jones, C.S., "Extracting factors from heteroskedastic asset returns," *Journal of Financial Economics*, 2001, *62*, 293–325.

Trzcinka, C., "On the Number of Factors in the Arbitrage Pricing Model," *Journal of Finance*, 1986, *41*, 347–368.

Most Recent Working Paper

NIPE WP 08/2020	Engle, R. F., and **Campos-Martins, S.**, Measuring and Hedging Geopolitical Risk, 2020
NIPE WP 07/2020	Boukari, M., and **Veiga, F. J.**, Fiscal Forecast Manipulations and Electoral Results: Evidence from Portuguese Municipalities, 2020
NIPE WP 06/2020	**Alexandre, F., Cruz, S**. and **Portela, M.**, Financial distress and the role of management in micro and small-sized firms, 2020
NIPE WP 05/2020	Cooke, D., Ana P. Fernandes and Priscila **Ferreira**, Entry Deregulation, Firm Organization and Wage Inequality, 2020
NIPE WP 04/2020	**Fernando Alexandre**, Pedro Bação, **João Cerejeira**, **Hélder Costa** and **Miguel Portela**, Minimum wage and financially distressed firms: another one bites the dust, 2020
NIPE WP 03/2020	**Luís Sá** and **Odd Rune Straume**, Quality provision in hospital markets with demand inertia: The role of patient expectations, 2020
NIPE WP 02/2020	**Rosa-Branca Esteves**, Liu Qihong and Shuai, J., Behavior-Based Price Discrimination with Non-Uniform Distribution of Consumer Preferences, 2020
NIPE WP 01/2020	Diogo Teixeira and **J. Cadima Ribeiro,** "Residents' perceptions of the tourism impacts on a mature destination: the case of Madeira Island", 2020
NIPE WP 17/2019	Liao, R. C., **Loureiro, G.**, and Taboada, A. G., "Women on Bank Boards: Evidence from Gender Quotas around the World", 2019
NIPE WP 16/2019	**Luís Sá,** "Hospital Competition Under Patient Inertia: Do Switching Costs Stimulate Quality Provision?", 2019
NIPE WP 15/2019	**João Martins** and **Linda G. Veiga**, "Undergraduate students' economic literacy, knowledge of the country's economic performance and opinions regarding appropriate economic policies", 2019
NIPE WP 14/2019	**Natália P. Monteiro, Odd Rune Straume** and **Marieta Valente,** "Does remote work improve or impair firm labour productivity? Longitudinal evidence from Portugal", 2019
NIPE WP 13/2019	**Luís Aguiar-Conraria,** Manuel M. F. Martins and **Maria Joana Soares**, " Okun's Law Across Time and Frequencies", 2019
NIPE WP 12/2019	Bohn, F., and **Veiga, F. J.,** "Political Budget Forecast Cycles", 2019
NIPE WP 11/2019	**Ojo, M. O., Aguiar-Conraria, L**. and **Soares, M. J., "**A Time-Frequency Analysis of Sovereign Debt Contagion in Europe", 2019
NIPE WP 10/2019	Lommerud, K. E., Meland, F. and **Straume, O. R.**, "International outsourcing and trade union (de-) centralization", 2019
NIPE WP 09/2019	**Carvalho, Margarita** and **João Cerejeira**, "Level Leverage decisions and manager characteristics",2019
NIPE WP 08/2019	**Carvalho, Margarita** and **João Cerejeira**, "Financialization, Corporate Governance and Employee Pay: A Firm Level Analysis", 2019
NIPE WP 07/2019	**Carvalho, Margarita** and **João Cerejeira,** "Mergers and Acquisitions and wage effects in the Portuguese banking sector", 2019
NIPE WP 06/2019	Bisceglia, Michele, Roberto Cellini, Luigi Siciliani and **Odd Rune Straume**, "Optimal dynamic volume-based price regulation", 2019
NIPE WP 05/2019	Hélia Costa and **Linda Veiga**, "Local labor impact of wind energy investment: an analysis of Portuguese municipalities", 2019
NIPE WP 04/2019	**Luís Aguiar-Conraria,** Manuel M. F. Martins and **Maria Joana Soares**, " The Phillips Curve at 60: time for time and frequency", 2019
NIPE WP 03/2019	**Luís Aguiar-Conraria,** Pedro C. Magalhães and Christoph A. Vanberg, "What are the best quorum rules? A Laboratory Investigation", 2019
NIPE WP 02/2019	**Ghandour, Ziad R.**, "Public-Private Competition in Regulated Markets", 2019
NIPE WP 01/2019	**Alexandre, Fernando**, Pedro Bação and **Miguel Portela**, "A flatter life-cycle consumption profile", 2019
NIPE WP 21/2018	**Veiga, Linda**, Georgios Efthyvoulou and Atsuyoshi Morozumi, "Political Budget Cycles: Conditioning Factors and New Evidence", 2018