INTERNATIONAL MONETARY FUND

Peering Through the Fog of Uncertainty: Out-of-Sample Forecasts of Post-Pandemic Tourism

Serhan Cevik

WP/23/70

2023
MAR

WORKING PAPER

 WP/23/70

IMF Working Paper

European Department

Peering Through the Fog of Uncertainty: Out-of-Sample Forecasts of Post-Pandemic Tourism

Prepared by Serhan Cevik[1]

Authorized for distribution by Bernardin Akitoby

March 2023

Abstract

This paper uses an augmented gravity model framework to investigate the historical impact of infectious diseases on international tourism and develops an out-of-sample prediction model. Using bilateral tourism flows among 38,184 pairs of countries during the period 1995–2017, I compare the forecasting performance of alternative specifications and estimation methods. These computations confirm the statistical and economic significance of infectious-disease episodes in forecasting international tourism flows. Including infectious diseases in the model improves forecast accuracy by an average of 4.5 percent and as much as 7 percent relative to the standard gravity model. The magnitude of these effects, however, is likely to be much greater in the case of COVID-19, which is a highly contagious virus that has spread fast throughout populations across the world.

JEL Classification Numbers:	C21; C23; F10; F11; F14; F47; R12
Keywords:	Infectious diseases; tourism flows; gravity model; predictability; out-of-sample forecasting

Author’s E-Mail Address:	scevik@imf.org

[1] The author would like to thank Alice Fan, Gyowon Gwon, Axel Schimmelpfennig, and Glebs Starovoits for their insightful comments and suggestions.

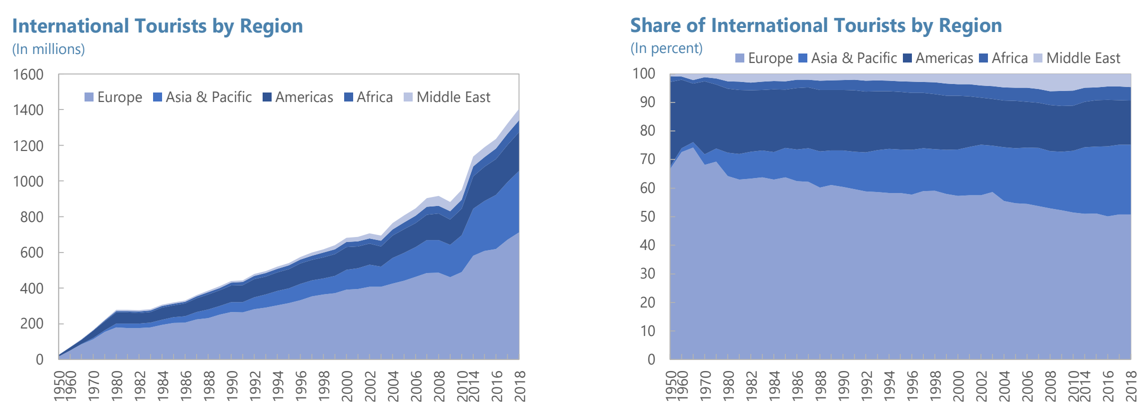
Prediction is very difficult, especially if it's about the future.
—Nils Bohr

I. Introduction

Over the past two years, the number of COVID-19 cases has reached 676.3 million, resulting more than 6.8 million deaths across the world and becoming a global crisis like no other in modern history.[2] The extensive containment and mitigation measures designed to restrict mobility and slow down the spread of COVID-19 caused the largest post-war recession (Coibon, Gorodnichenko, Weber, 2020; Eichenbaum, Rebelo, and Trabandt, 2020; Fornaro and Wolf, 2020; Hassan and others, 2020; Ludvigson, Ma, and Ng, 2020; Cevik and Miryugin, 2021). At the same time, governments implemented policy measures to cushion the consequences of the pandemic and stimulate economic recovery. As a result, the COVID-19 pandemic—similar to other episodes of infectious diseases in history—has had largely heterogeneous effects across the world, reflecting the varying degree of exposure to the virus, composition of economic sectors, overall level of preparedness, and capacity for adequate public response.

Tourism is one of the most affected sectors, due to containment and mitigation measures aimed to slow the spread of the virus. According to the United Nations World Tourism Organization (UNWTO), international tourist arrivals declined by 75 percent in 2020 and 70 percent in 2021, compared to 2019. This is an unprecedented shock across the world, but especially in tourism-dependent economies. One of the most important engines of economic growth before the COVID-19 pandemic, tourism accounted for more than 10 percent of the global economy and a large share of total employment and export earnings in many countries. Furthermore, the tourism sector is closely interconnected to others in the economy, including accommodation and dining, retail and marketing, and transportation and aviation, forming increasingly complex supply chain (Goretti and others, 2021). Even though past pandemics were generally short-lived,

International Tourism Flows



Source: UNWTO; author's calculations.

[2] The latest figures can be found at John Hopkins University's Center for Systems Science and Engineering: https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6.

how each country comes out of the current crisis depends on policy choices made during the pandemic, the required adjustment, and the economic and institutional strength prior to COVID-19. Therefore, analyzing the evolution of international tourism flows during past epidemics can shed more light on the impact of COVID-19 and improve the prediction of recovery trajectories as the world moves into a new phase of the pandemic.

Past infectious-disease episodes are shown to have a significant effect on bilateral tourism flows across the world, with greater impact in Asia, Latin America, and the Caribbean (Cevik 2022). Statistical significance, however, cannot substitute for out-of-sample predictive power, which can tell more about the external validity of the empirical results. Accordingly, using an augmented gravity framework, this paper develops an out-of-sample prediction model for international tourism flows among 38,184 pairs of countries during the period 1995–2017 and tests the forecasting performance of alternative specifications and estimation methods. These computations confirm the significance of infectious-disease episodes in forecasting international tourism flows. I find that including infectious diseases in the model improves forecast accuracy by an average of 4.5 percent and as much as 7 percent relative to the standard gravity model. The magnitude of these effects, however, is likely to be much greater in the case of COVID-19, which is a highly contagious virus that has spread throughout populations around the globe. Furthermore, the augmented forecasting model presented in this paper would help analyze growth dynamics during and after the next significant episode of infectious diseases

The remainder of this paper is structured as follows. Section II provides an overview of the data used in the empirical analysis. Section III describes the econometric methodology. Section IV presents the empirical results, including a series of robustness checks. Finally, Section V summarizes and provides concluding remarks.

II. Data Overview

The empirical analysis presented in this study is based on an unbalanced panel of annual observations for 38,184 pairs of countries during the period 1995–2017.[3] Bilateral tourism flows for 172 countries of origin and 222 countries of destination are taken from the UNWTO database, yielding a dataset of over 261,488 observations over the sample period. The main economic variable in the gravity model is the size as measured by the level of real GDP in origin and destination countries. I also include population and life expectancy in origin and destination countries as additional control variables to better capture the size effect and overall health conditions. These series are obtained from the World Bank's World Development Indicators (WDI) database. The key explanatory variable of interest in this paper is the number of confirmed infectious-disease cases, including Ebola, malaria, SARS, and yellow fever, which is obtained from the World Health Organization (WHO) database.

Standard gravity variables such as bilateral distance between countries, common official language, colonial history and geographical contiguity are taken from the Centre d'Etudes

[3] The list of countries, including territories, is presented in Appendix Table A1.

Prospectives et d'Informations Internationales (CEPII) database, as presented in Mayer and Zignago (2011) and Conte, Cotteriaz, and Mayer (2022). Geographic distance is measured as the great-circle distance in kilometers between the capital cities of each country pair. Traditionally, distance in the gravity model is not just a measure of geographical distance, but it also reflects transportation costs and other trade barriers. Binary variables for language, colonial history and geographical contiguity are assigned a value of 1 if a country pair share a common official language, a colonial tie, and an adjacent border and a value of 0 otherwise.

Descriptive statistics for the variables used in the empirical analysis are presented in Table 1. There is a significant degree of dispersion across countries in terms of international tourist flows and considerable heterogeneity in the occurrence of infectious diseases. It is essential to analyze the time-series properties of the data to avoid spurious results by conducting panel unit root tests. Accordingly, the stationarity of all variables is checked by applying the Im-Pesaran-Shin (2003) procedure, which is widely used in the empirical literature to conduct a panel unit root test. The results, available upon request, indicate that the variables used in the analysis are stationary after logarithmic transformation.

Table 1. Descriptive Statistics

Variables	Obs.	Mean	Std. Dev.	Min.	Max.
International tourism flows	261,488	78,737	918,199	1	81,100,000
Real GDP					
Origin countries	410,680	13,155	16,243	184	111,968
Destination countries	381,210	16,926	22,269	184	194,188
REER, destination	371,190	100.5	25.5	7.4	740.6
Distance	370,208	7,273	4,564	27	19,951
Common language	370,208	0.173	0.378	0.000	1.000
Colonial history	370,208	0.015	0.123	0.000	1.000
Geographical contiguity	370,208	0.027	0.163	0.000	1.000
Life expectancy, destination	382,122	70.4	9.4	31.0	85.4
Population					
Origin countries	411,470	54,300,000	186,000,000	9,298	1,390,000,000
Destination countries	395,038	45,900,000	160,000,000	9,298	1,400,000,000
Infectious diseases					
Origin countries					
Ebola	411,470	22	496	0	14,124
Malaria	411,470	69,328	631,885	0	15,000,000
SARS	411,470	3	115	0	5,327
Yellow fever	411,470	3	40	0	1,192
Destination countries					
Ebola	386,101	15	413	0	14,124
Malaria	386,101	130,202	746,793	0	15,000,000
SARS	386,101	3	108	0	5,327
Yellow fever	378,948	3	34	0	1,192

Source: UNWTO; IMF; World Bank; WHO; author's calculations.

III. Empirical Methodology

The gravity framework is widely used in the literature to analyze the patterns of international trade and capital movements, as well as migration and tourism flows (Anderson, 1979; Anderson and van Wincoop, 2003; Bergstrand and Egger, 2007; Gil-Pareja, Llorca-Vivero, and Martínez-Serrano, 2007; Head and Ries, 2008; Santana-Gallego, Ledesma-Rodríguez, and Pérez-Rodríguez, 2010). But there is scarce research on modeling bilateral tourist movements in a gravity framework and taking into account infectious diseases. Most studies in this context look at the impact of disease outbreaks, such as the SARS and avian flu epidemics, on tourism in a specific country or region over a short period of time (Zeng, Carter, and De Lacey, 2005; Cooper, 2006; Wilder-Smith, 2006; Kuo and others, 2008). Using dummy variables infectious diseases, Roselló, Santana-Gallego, and Awan (2017) show that the eradication of infectious diseases benefits countries in terms of tourism flows and revenues. More recently, taking advantage of a rich dataset covering 38,184 pairs of countries over the period 1995–2017, Cevik (2022) finds strong evidence that international tourism is adversely affected by the risk of infectious diseases and the magnitude of this negative effect is statistically and economically significant.

The gravity model developed in this paper borrows a number of insights from structural gravity models pioneered in the international-trade literature. Bilateral flows between two countries are proportionate to economic size and inversely proportionate to geographic distance:

$$T_{ij} = \mathrm{B}\frac{(GDP_i)^{\alpha}(GDP_j)^{\gamma}}{(Dist_{ij})^{\vartheta}}\, U_{ij} \qquad (1)$$

where T_{ij} denotes international tourist flows between countries i (origin) and j (destination); GDP refers to the gross domestic product of each country; $Dist_{ij}$ is the distance between countries i and j; and $U_{ij}$is a log-normal distributed error term. In other words, the number of visitors between two countries depends on the economic sizes of the countries and the distance between them. In a panel data context, however, this expression can be transformed using natural logarithms to:

$$ln\,(T_{ijt}) = \beta + \alpha\, ln\,(GDP_{it}) + \gamma\, ln\,(GDP_{jt}) + \vartheta\, ln\,(Dist_{ij}) + \eta_i + \varphi_j + \mu_t + \varepsilon_{ijt} \qquad (2)$$

in which T_{ijt} denotes international tourist flows between countries i (origin) and j (destination) at time t; GDP is the level of per capita income in origin and destination country, respectively, at time t; $Dist_{ij}$ is the physical distance between countries i (origin) and j (destination); the η_i, φ_j and μ_t coefficients designate the country fixed effects capturing all time-invariant factors in origin and destination country and the time fixed effects controlling for common shocks that may affect international tourism across all countries in a given year, respectively. ε_{ijt} is an idiosyncratic error term that meets the standard assumptions of zero mean and constant variance.

Since the objective is to understand the effect of infectious diseases on international tourism, the parsimonious gravity model is augmented with additional control variables along with the number of confirmed infectious-disease cases:

$$ln\,(T_{ijt}) = \beta + \alpha ln\,(GDP_{it}) + \gamma ln\,(GDP_{jt}) + \vartheta ln\,(Dist_{ij}) + \delta X_{ijt} + \varphi \ln\left(Vir_{ijt}\right) + \eta_i + \varphi_j + \mu_t + \varepsilon_{ijt} \quad (3)$$

where X_{ij} denotes a vector of control variables, including the REER, population, life expectancy, and binary variables for common language, colonial history and geographical contiguity; Vir_{ijt} denotes the number of confirmed cases of Ebola, malaria, SARS, and yellow fever scaled by population in origin and destination countries. To account for possible heteroskedasticity, robust standard errors are clustered at the country-pair level.[4]

Most gravity models are estimated with cross-sectional data, which may lead to biased results due to potential correlation between explanatory variables and unobservable country characteristics as it does not control for heterogeneity. Panel data estimations help address such econometric concerns by controlling for country and time fixed effects (Egger, 2000). Therefore, in this paper, I estimate the gravity model with the Poisson pseudo maximum likelihood estimation (PPML) procedure recommended by Santos Silva and Tenreyro (2006) and the two-stage least squares with instrumental variable (2SLS-IV) methodology using the lagged infectious disease as instrument to account for potential endogeneity.[5]

I compare the out-of-sample forecasting performance of alternative gravity models of bilateral tourism flows by partitioning the original sample period (1995–2017) into two subsamples: (1) the estimation sample (1995–2014) and the forecasting sample (2015–2017). In other words, to predict bilateral tourism flows during 2015–2017, I utilize the data covering the full sample of countries over the period 1995–2014 to estimate each model. Then, I compared the predicted bilateral tourism flows with the actual data to compute the forecast errors. For this exercise, I focus only on Ebola and SARS—infectious diseases similar to COVID-19 with human-to-human transmission—and obtain the coefficient estimates for each specification.

To evaluate forecast accuracy of these alternative models, I employ the mean absolute error (MAE), the root mean squared error (RMSE) and the Theil Inequality Coefficient (U-Theil), which are the most commonly used metrics in the literature and defined by the following equations:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|\hat{A}_{t,c} - A_{t,c}\right| \quad (4)$$

[4] The results remain broadly unchanged when standard errors are clustered at the country level.

[5] The gravity model is also estimated using the OLS as a further robustness check, which are available upon request. Since the objective is to include standard time-invariant gravity factors (distance, common language, colonial history, geographical contiguity) in the panel regressions, the OLS model is estimated via the random-effects regression, instead of the fixed-effects model that would remove time-invariant variables. However, the fixed-effects estimations with origin and destination fixed effects controlling for all possible time-invariant country characteristics yield similar results.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{A}_{t,c} - A_{t,c})^2} \quad (5)$$

$$U - Theil = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{A}_{t,c} - A_{t,c})^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{A}_{t,c})^2} + \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_{t,c})^2}} \quad (6)$$

in which $\hat{A}_{t,c}$ and $A_{t,c}$ are the predicted and actual bilateral tourism flows at time *t*, respectively, and *n* is the number of observations in the sample. The model with the lowest MAE, RMSE, and U-Theil values is considered to better forecast accuracy.

IV. Estimation Results

A. Panel Estimations

The baseline gravity model described in Equation (3) is estimated using the PPML method for and start with a specification including only macroeconomic and demographic variables and standard gravity factors in column [1] of Table 2 as a point of reference. The number of infectious diseases is then introduced into the regression in column [2] for Ebola, column [3] for malaria, column [4] for SARS, and column [5] for yellow fever. The results demonstrate a consistent picture with the signs of all estimated parameters corresponding to their expected values across different specifications. I estimate the gravity model using the 2SLS-IV method and obtain similar results, albeit with a considerable increase in the magnitude and statistical significance of the coefficients on infectious diseases (Table 3).[6]

Standard gravity indicators. With regards to the standard gravity variables, I find that the level of income in both origin and destination countries have a positive impact on bilateral tourism flows, suggesting that international tourism is significantly related to the two countries' economic size. Distance between the countries, on the other hand, is negatively associated with bilateral tourism flows, representing an obstacle for international travel as expected. The greater the distance between the two countries, the smaller the flow of tourists across the two countries, due to higher cost of travel. This is also consistent with the positive effect of the geographical contiguity variable, indicating that tourists tend to travel more to closer destinations. Cultural similarities and historical ties, proxied by common official language and colonial relations, are found to have the expected positive effects on bilateral tourism flows. Likewise, demographic factors, measured by population in origin and destination countries, also contribute to stronger tourism flows between country pairs across the world.

[6] The 2SLS-IV estimation increases the magnitude of the cumulative coefficient on the infectious disease variable (SARS) to -0.465 compared to -0.043 in the baseline estimation, strongly supporting the contemporaneous impact of infectious diseases on international travel both in terms of magnitude and statistical significance.

Table 2. Infectious Diseases and International Tourism—PPML Estimations

(Dependent variable: Bilateral tourism flows)	[1]	[2]	[3]	[4]	[5]
Real GDP, origin	0.129***	0.127***	0.127***	0.127***	0.131***
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
Real GDP, destination	0.129***	0.129***	0.129***	0.129***	0.132***
	[0.007]	[0.008]	[0.008]	[0.008]	[0.008]
Distance	-0.228***	-0.232***	-0.231***	-0.232***	-0.229***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Common language	0.176***	0.174***	0.174***	0.174***	0.180***
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Colonial history	0.080***	0.083***	0.083***	0.083**	0.076***
	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]
Geographical contiguity	0.034	0.035	0.035	0.035	0.031
	[0.019]	[0.016]	[0.016]	[0.016]	[0.016]
Population, origin	0.063***	0.065***	0.059***	0.063***	0.066***
	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Population, destination	0.080***	0.078***	0.082***	0.078***	0.084***
	[0.012]	[0.012]	[0.012]	[0.013]	[0.012]
REER, destination	0.000***	0.000***	0.000***	0.000***	0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Life expectancy, destination	0.097	0.103	0.113	0.091	0.088
	[0.043]	[0.043]	[0.044]	[0.044]	[0.043]
Ebola					
Origin		-0.013***			
		[0.001]			
Destination		-0.010***			
		[0.002]			
Malaria					
Origin			0.001		
			[0.000]		
Destination			-0.001		
			[0.000]		
SARS					
Origin				-0.003***	
				[0.02]	
Destination				-0.040***	
				[0.02]	
Yellow fever					
Origin					0.003
					[0.000]
Destination					-0.001
					[0.001]
Number of observations	224,019	215,589	215,589	215,589	219,132
Origin FE	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.77	0.84	0.80	0.83	0.80

Note: The dependent variable is bilateral tourism flows (in log form). Robust standard errors, clustered at the country level, are reported in brackets. A constant is included in each regression, but not shown in the table. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 3. Infectious Diseases and International Tourism—2SLS-IV Estimations

(Dependent variable: Bilateral tourism flows)				
	[1]	[2]	[3]	[4]
Real GDP, origin	0.933***	0.935***	0.959***	0.957***
	[0.012]	[0.012]	[0.014]	[0.012]
Real GDP, destination	0.865***	0.856***	0.861***	0.896***
	[0.013]	[0.013]	[0.015]	[0.013]
Distance	-1.717***	-1.717***	-1.716***	-1.717***
	[0.015]	[0.015]	[0.015]	[0.015]
Common language	1.241***	1.241***	1.241***	1.272***
	[0.034]	[0.034]	[0.034]	[0.034]
Colonial history	0.849***	0.850***	0.850***	0.832***
	[0.086]	[0.086]	[0.086]	[0.084]
Geographical contiguity	1.210***	1.207***	1.208***	1.163***
	[0.067]	[0.067]	[0.067]	[0.066]
Population, origin	0.523***	0.517***	0.514***	0.498***
	[0.022]	[0.023]	[0.023]	[0.022]
Population, destination	0.526***	0.562***	0.514***	0.576***
	[0.023]	[0.024]	[0.024]	[0.022]
REER, destination	0.001***	0.001***	0.001***	0.001***
	[0.000]	[0.000]	[0.000]	[0.000]
Life expectancy, destination	0.097	0.039	0.214	0.209
	[0.070]	[0.078]	[0.081]	[0.068]
Ebola				
Origin	-0.065***			
	[0.005]			
Destination	-0.089***			
	[0.008]			
Malaria				
Origin		-0.001		
		[0.001]		
Destination		-0.007		
		[0.001]		
SARS				
Origin			-0.387***	
			[0.092]	
Destination			-0.078***	
			[0.104]	
Yellow fever				
Origin				0.004
				[0.012]
Destination				-0.017
				[0.104]
Number of observations	210,221	210,221	210,221	213,645
Origin FE	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Pseudo R^2	0.83	0.83	0.83	0.84

Note: The dependent variable is bilateral tourism flows (in log form). Robust standard errors, clustered at the country level, are reported in brackets. A constant is included in each regression, but not shown in the table. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Role of infectious diseases. With regards to the main explanatory variable of interest, estimation results establish a significant effect of infectious-disease episodes on international tourism flows, but with variation in magnitude and statistical significance depending on the nature of the disease. The estimated coefficients on malaria and yellow fewer are considerably smaller in magnitude, whereas the coefficients on Ebola and SARS are found to be both statistically and economically significant. These results are robust to alternative estimations and specifications, including after controlling for health infrastructure. In the case of SARS, for example, a 10 percent increase in the number of confirmed cases leads, on average, to a reduction of 4.3 percent in international tourist flows. There is, however, significant heterogeneity across country groups in the impact of pandemics on tourism, depending on the level of income and health infrastructure.

Every infectious disease is different in important ways, but there are significant similarities between SARS and COVID-19, which belong to the same family of coronavirus. Scaling the estimated coefficient of SARS to the prevalence of COVID-19 as measured by the number of confirmed cases in population would yield an approximate decline of 82.5 percent in international tourism flows, which is consistent with the actual drop of 75 percent in 2020 and 70 percent in 2021 compared to the pre-pandemic level. The estimated differences in how infectious diseases affect international tourism flows likely reflect disease-specific characteristics.

Vector of transmission. Malaria and yellow fever are transmitted by mosquitoes, but Ebola and SARS—similarly to COVID-19—are spread from human to human. Accordingly, while malaria and yellow fever may be endemic in rural areas, Ebola and SARS could spread more easily in densely populated cities and airports.

Existence of treatment or vaccine. Although a vaccine for yellow fever and treatments for malaria exist, to the authors' knowledge there is no such treatment or vaccine against Ebola or SARS. Consequently, infection risks of these diseases have a greater effect on international tourism flows, especially to countries with weak health infrastructure.

Temporary outbreak vs. endemic presence. When a disease is endemic like malaria and yellow fever, there is no point in delaying travel as long as precautions can be taken. Outbreaks of Ebola and SARS, on the other hand, are temporary in nature and incentivize tourists to delay visiting a particular country until the outbreak is over.

Partitioning the sample into income groups and geographical regions highlights heterogeneity on how the risk of infectious diseases affects tourism flows. These estimation results, presented in Table 5, show a substantial contrast between advanced economies and developing countries. Although infectious diseases appear to have statistically insignificant effect on tourism flows to advanced economies, the magnitude and statistical significance of the impact of infectious diseases are much greater in developing countries, where such diseases tend to be more prevalent and health infrastructure lags behind.[7] For example, in the case of SARS, a 10 percent

[7] Unlike past episodes, however, the impact of COVID-19 on tourism flows will be similar across all country groups, given the extent of containment measures put in place by all countries regardless the level of income.

increase in the number of infections leads to a decline of 3.2 percent in bilateral tourism flows in advanced economies, but almost 12 percent in developing countries. These findings show systemic differences among geographical regions: the disease impact on international tourism flows is significantly greater in Asia, Latin America, and the Caribbean than the rest of the world.

Table 4. Infectious Diseases and Tourism—Robustness Checks

(Dependent variable: Bilateral tourism flows)			
	Truncated sample	Sub-sample (1995-2007)	Additional controls
Real GDP, origin	0.961***	1.259***	1.023***
	[0.013]	[0.028]	[0.018]
Real GDP, destination	0.808***	0.751***	0.967***
	[0.015]	[0.028]	[0.019]
Distance	-1.632***	-1.688***	-1.704**
	[0.015]	[0.021]	[0.018]
Common language	1.202***	1.170***	1.233***
	[0.033]	[0.046]	[0.039]
Colonial history	0.755***	0.919***	0.864***
	[0.089]	[0.106]	[0.093]
Geographical contiguity	1.116***	1.232***	1.261**
	[0.072]	[0.084]	[0.074]
Population, origin	0.478***	1.224***	0.554***
	[0.022]	[0.057]	[0.033]
Population, destination	0.478**	0.305***	0.625***
	[0.023]	[0.057]	[0.032]
REER, destination	0.001**	0.001**	0.001**
	[0.000]	[0.000]	[0.000]
Life expectancy, destination	0.151	0.060	0.711***
	[0.078]	[0.136]	[0.194]
Hospital beds, destination			0.027***
			[0.010]
SARS			
Origin	-0.224***	-0.281***	-0.224***
	[0.085]	[0.041]	[0.095]
Destination	-0.028***	-0.016***	-0.011***
	[0.091]	[0.050]	[0.103]
Number of observations	176,489	96,416	111,591
Origin FE	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Adjusted R^2	0.79	0.83	0.83

Note: The dependent variable is bilateral tourism flows (in log form). Robust standard errors, clustered at the country level, are reported in brackets. A constant is included in each regression, but not shown in the table. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

As a further check for the robustness of the results, I estimate the gravity model with alternative specifications.[8] First, the sample is truncated at the 5th and 95th percentiles to remove the potential impact of extreme outliers. Second, the gravity model is estimated for a sub-sample of 1995-2007 to exclude the period after the global financial crisis. Third, additional health-related variables—life expectancy and the number of hospital beds per 1,000 people—are introduced to address omitted-variable bias and capture the impact of health conditions and infrastructure. These results, presented in Table 4, show that the negative and economically significant relationship between infectious diseases and international tourism flows remains broadly unchanged. Estimating the model with the truncated sample and for the period excluding the global financial crisis yields higher coefficients on the infectious-disease variable. Adding health variables into the regression model reveals that health conditions and infrastructure in destination countries matter for bilateral tourism flows.

Table 5. Infectious Diseases and Tourism—Estimations by Income Group and Region

(Dependent variable: Bilateral tourism flows)	Advanced	Developing	Africa	Asia	Europe	Latin America	Middle East
Real GDP, origin	1.008***	0.935***	0.860***	1.044***	0.754***	0.889***	1.169***
	[0.017]	[0.020]	[0.044]	[0.054]	[0.044]	[0.035]	[0.052]
Real GDP, destination	1.083***	0.744***	0.304***	1.201***	0.575***	0.744***	0.371***
	[0.044]	[0.024]	[0.040]	[0.077]	[0.045]	[0.047]	[0.060]
Distance	-1.351***	-1.821***	-1.512**	-1.932***	-1.518***	-1.737***	-1.482***
	[0.026]	[0.019]	[0.053]	[0.077]	[0.085]	[0.051]	[0.072]
Common language	0.584***	1.404***	1.170***	0.697***	0.863	1.451***	0.781***
	[0.054]	[0.040]	[0.058]	[0.112]	[0.406]	[0.064]	[0.126]
Colonial history	1.164***	0.848***	0.231	1.116**	0.109	0.541	0.526
	[0.082]	[0.146]	[0.499]	[0.409]	[0.208]	[0.529]	[0.388]
Geographical contiguity	0.480***	1.307***	1.126**	0.936***	1.525***	0.970***	1.662***
	[0.106]	[0.079]	[0.148]	[0.185]	[0.167]	[0.150]	[0.198]
Population, origin	0.219***	0.660***	1.010***	0.824***	0.860***	0.797***	0.156
	[0.030]	[0.031]	[0.065]	[0.084]	[0.068]	[0.060]	[0.084]
Population, destination	0.231***	0.433***	0.398***	0.393	0.316	0.729***	0.885***
	[0.063]	[0.031]	[0.133]	[0.179]	[0.148]	[0.114]	[0.065]
REER, destination	0.006***	-0.000***	0.000	-0.002***	0.000	0.000***	0.000
	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]
Life expectancy, destination	5.847***	0.585***	0.366*	1.668**	3.911***	0.745	0.887
	[0.528]	[0.087]	[0.146]	[0.554]	[1.264]	[0.420]	[0.659]
SARS							
Origin	-0.472	-0.556***	-0.362	-0.993***	-0.144	-0.455**	-0.850*
	[0.136]	[0.120]	[0.221]	[0.315]	[0.322]	[0.219]	[0.331]
Destination	-0.246	-0.243***	-0.207	-0.311***	-0.761	-0.204	-0.229
	[0.110]	[0.091]	[0.544]	[0.191]	[1.505]	[0.630]	[0.371]
Number of observations	70,721	139,500	36,232	23,922	23,794	33,750	21,802
Origin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.88	0.85	0.81	0.86	0.85	0.85	0.81

Note: The dependent variable is bilateral tourism flows (in log form). Robust standard errors, clustered at the country level, are reported in brackets. A constant is included in each regression, but not shown in the table. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

[8] To exhibit a concise table, the robustness checks are presented for only SARS, but the results remain consistent for other infectious diseases.

B. Out-of-Sample Prediction

Given the nature of the question central to this paper, however, statistical significance cannot substitute for out-of-sample predictive power, which can tell more about the external validity of the empirical results. Therefore, one of the main objectives of this paper is to formally assess the out-of-sample forecast performance of alternative gravity models and estimation methodologies in predicting bilateral tourism flows, conditionally on the information available up to and including 2014. I estimate and present three evaluation criteria to judge the forecasting performance of the models—MAE, RMSE and U-Theil, which are widely used in the literature and yield similar results.

These computations, presented in Table 6, confirm the statistical and economic significance of infectious-disease episodes in several out-of-sample forecasting exercises. The inclusion of infectious diseases in the model lowers the RMSE of bilateral tourism flow forecasts by an average of 4.5 percent relative to the standard gravity model without the number of infectious-disease cases. This means greater precision in forecasting international tourism flows when the model is augmented with the information on infectious diseases.

Estimating the augmented gravity model with the 2SLS-IV approach is able to reduce the RMSE even—by as much as 7 percent in the case of SARS, which indicates a higher degree of predictability. In other words, augmenting the model with infectious diseases improves forecast accuracy, corroborating that information on past infectious-diseases episodes is valuable in predicting bilateral tourism flows in the future. Other metrics (MAE and U-Theil) used to evaluate forecast accuracy yield similar results showing that the augmented gravity models improve upon

Table 6. Out-of-Sample Forecast Performance

(Dependent variable: Bilateral tourism flows)			
	Standard	Ebola	SARS
PPML models			
MAE	5.103	5.086	**4.954**
RMSE	5.806	5.791	**5.675**
U-Theil	0.625	0.624	**0.613**
2SLS-IV models			
MAE	1.168	1.165	**1.096**
RMSE	1.675	1.652	**1.561**
U-Theil	0.118	0.116	**0.111**

Note: Each model is trained with the data covering the period 1995-2014, then tested in forecasting on the period 2015-2017. The model with the lowest MAE, RMSE, and U-Theil values is considered to better forecast accuracy, which is shown in bold.

the standard model without infectious-disease episodes.[9] Overall, these estimates should be considered as a lower bound in assessing the contribution of infectious diseases to forecasting. In the case of COVID-19, the magnitude of these effects is likely to be much greater due to widespread containment measures that led to an unprecedented collapse in international tourism flows.

V. Conclusion

The COVID-19 pandemic is considerably different than past epidemics in modern times, but past epidemics can still shed light on its impact on international travel and tourism. Using a rich dataset of 38,184 pairs of countries over the period 1995-2017 and an augmented gravity model, the empirical analysis shows that bilateral tourism flows are across the world are adversely affected by the risk of infectious disease, and the magnitude of this negative effect is statistically and economically significant. In the case of SARS, for example, I find that a 10 percent increase in the number of confirmed cases leads, on average, to a reduction of 4.3 percent in international tourist arrivals. Furthermore, while infectious diseases appear to have a smaller and statistically insignificant negative effect on tourism flows to advanced economies, the magnitude and statistical significance of the impact of infectious diseases are much greater in developing countries, where such diseases tend to be more prevalent and health infrastructure lags behind.

Scaling the estimated coefficient of SARS to the prevalence of COVID-19 would yield an approximate decline of 82.5 percent in international tourism flows, which is broadly consistent with the actual drop of 75 percent in 2020 and 70 percent in 2021 compared to the pre-pandemic level. This shows that the information on infectious diseases should not be ignored in forecasting international tourism flows across the world. Accordingly, I develop an out-of-sample prediction model for bilateral tourism flows among 38,184 pairs of countries and test the forecasting performance of alternative specifications and estimation methods. These computations confirm the significance of infectious-disease episodes in forecasting international tourism flows. Including infectious diseases in the model improves forecast accuracy by an average of 4.5 percent and as much as 7 percent relative to the standard gravity model.

These estimates, however, should be treated as a lower bound. In the case of COVID-19, the magnitude of these effects is likely to be much greater, as policymakers have imposed strict containment and lockdown measures in order to contain the spread of the highly contagious coronavirus that led to an unprecedented collapse in international tourism flows.[10] Furthermore, the augmented forecasting model presented in this paper would help analyze growth dynamics during and after the next significant episode of infectious diseases.

---

[9] I also estimate these models separately for subsamples of countries in Asia and Latin America and the Caribbean and obtain similar forecast errors.

[10] International tourist arrivals in 2022 remained 37 percent below the 2019 level, after declining by 70 percent in 2021 and 75 percent in 2020.

Appendix Table A1. List of Countries and Territories

Afghanistan	Denmark	Liberia	Rwanda
Albania	Djibouti	Libya	Saba
Algeria	Dominica	Liechtenstein	Saint Eustatius
American Samoa	Dominican Republic	Lithuania	Saint Maarten
Andorra	Ecuador	Luxembourg	Samoa
Angola	Egypt	Macao SAR	San Marino
Anguilla	El Salvador	Madagascar	Sao Tome And Principe
Antigua And Barbuda	Equatorial Guinea	Malawi	Saudi Arabia
Argentina	Eritrea	Malaysia	Senegal
Armenia	Estonia	Maldives	Serbia
Aruba	Eswatini	Mali	Seychelles
Australia	Ethiopia	Malta	Sierra Leone
Austria	Fiji	Marshall Islands	Singapore
Azerbaijan	Finland	Martinique	Slovak Republic
Bahamas, The	France	Mauritania	Slovenia
Bahrain	French Guiana	Mauritius	Solomon Islands
Bangladesh	French Polynesia	Mexico	Somalia
Barbados	Gabon	Micronesia	South Africa
Belarus	Gambia, the	Moldova	South Sudan
Belgium	Georgia	Monaco	Spain
Belize	Germany	Mongolia	Sri Lanka
Benin	Ghana	Montenegro	St. Kitts and Nevis
Bermuda	Greece	Montserrat	St. Lucia
Bhutan	Grenada	Morocco	St. Vincent and the Grenadines
Bolivia	Guadeloupe	Mozambique	Sudan
Bonaire	Guam	Myanmar	Suriname
Bosnia And Herzegovina	Guatemala	Namibia	Sweden
Botswana	Guinea	Nauru	Switzerland
Brazil	Guinea-Bissau	Nepal	Syria
British Virgin Islands	Guyana	Netherlands	Taiwan Province of China
Brunei Darussalam	Haiti	New Caledonia	Tajikistan
Bulgaria	Honduras	New Zealand	Tanzania
Burkina Faso	Hong Kong SAR	Nicaragua	Thailand
Burundi	Hungary	Niger	Timor-Leste
Cabo Verde	Iceland	Nigeria	Togo
Cambodia	India	Niue	Tonga
Cameroon	Indonesia	North Korea	Trinidad And Tobago
Canada	Iran	North Macedonia	Tunisia
Cayman Islands	Iraq	Northern Mariana Islands	Turkey
Central African Republic	Ireland	Norway	Turkmenistan
Chad	Israel	Oman	Turks And Caicos Islands
Chile	Italy	Pakistan	Tuvalu
China	Jamaica	Palau	Uganda
Colombia	Japan	Palestine	Ukraine
Comoros	Jordan	Panama	United Arab Emirates
Congo, Republic of	Kazakhstan	Papua New Guinea	United Kingdom
Cook Islands	Kenya	Paraguay	United States
Costa Rica	Kiribati	Peru	United States Virgin Islands
Côte d'Ivoire	Korea	Philippines	Uruguay
Croatia	Kuwait	Poland	Uzbekistan
Cuba	Kyrgyz Republic	Portugal	Vanuatu
Curacao	Lao P.D.R.	Puerto Rico	Venezuela
Cyprus	Latvia	Qatar	Vietnam
Czech Republic	Lebanon	Reunion	Yemen
Democratic Republic Of The Congo	Lesotho	Romania	Zambia
		Russia	Zimbabwe

References

Anderson, J. (1979). "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, Vol. 69, pp. 106-116.

Anderson, J., and E. van Wincoop (2003). "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, Vol. 93, pp. 170–192.

Bergstrand, J., and P. Egger (2007). "A Knowledge-and-Physical-Capital Model of International Trade, Foreign Direct Investment, and Foreign Affiliates Sales: Developed Countries," *Journal of International Economics*, Vol. 73, pp. 278–308.

Cevik, S. (2022). "Going Viral: A Gravity Model of Infectious Diseases and Tourism Flows," *Open Economies Review*, Vol. 33, pp. 141–156.

Cevik, S., and F. Miryugin (2021). "Pandemics and Firms: Drawing Lessons from History," *International Finance*, Vol. 24, pp. 276–297.

Coibon, O., Y. Gorodnichenko, and M. Weber (2020). "The Cost of the COVID-19 Crisis: Lockdowns, Macroeconomic Expectations, and Consumer Spending," NBER Working Paper No. 27141 (Cambridge, MA: National Bureau of Economic Research).

Conte, M., P. Cotteriaz, and T. Mayer (2022). "The CEPII Gravity Database," CEPII Working Paper No. 2022-05 (Paris: Centre d'Etudes Prospectives et d'Informations Internationales).

Cooper, M. (2006). "Japanese Tourism and the SARS Epidemic of 2003," *Journal of Travel and Tourism Marketing*, Vol. 19, pp. 117–131.

Egger, P. (2000). "A Note on the Proper Econometric Specification of the Gravity Equation," *Economic Letters*, Vol. 66, pp. 25–31.

Eichenbaum, M., S. Rebelo, and M. Trabandt (2020). "The Macroeconomics of Epidemics," NBER Working Paper No. 26882 (Cambridge, MA: National Bureau of Economic Research).

Fornaro, L., and M. Wolf (2020). "Covid-19 Coronavirus and Macroeconomic Policy," CEPR Discussion Paper No. 14529 (London: Centre for Economic Policy Research).

Gil-Pareja, S., R. Llorca-Vivero, and J. Martínez-Serrano (2007). "The Effect of EMU on Tourism," *Review of International Economics*, Vol. 15, pp. 302–312.

Goretti, M., L. Leigh, S. Cevik, S. Kaendera, D. Muir, S. Nadeem, and G. Salinas (2021). "Tourism in the Post-Pandemic World: Economic Challenges and Opportunities for Asia-Pacific and the Western Hemisphere," IMF Departmental Paper No. 21/02 (Washington, DC: International Monetary Fund).

Hassan, T., S. Hollander, L. van Lent, and A. Tahoun (2020). "Firm-Level Exposure to Epidemic Diseases: Covid-19, SARS, and H1N1," NBER Working Paper No. 26971 (Cambridge, MA: National Bureau of Economic Research).

Head, K., and J. Ries (2008). "FDI as an Outcome of the Market for Corporate Control: Theory and Evidence," *Journal of International Economics*, Vol. 74, pp. 2–20.

Im, K., M. Pesaran, and Y. Shin (2003). "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, Vol. 115, pp. 53–74.

Kuo, H-I., C-C. Chen, W-C. Tseng, L-F Ju, and B-W. Huang (2008). "Assessing Impacts of SARS and Avian Flu on International Tourism Demand to Asia," *Tourism Management*, Vol. 29, pp. 917–928.

Ludvigson, S., S. Ma, and S. Ng (2020). "COVID-19 and the Macroeconomic Effects of Costly Disasters," NBER Working Paper No. 26987 (Cambridge, MA: National Bureau of Economic Research).

Mayer, T., and S. Zignago (2011). "Notes on CEPII's Distance Measures: The GeoDist Database," CEPII Working Paper No. 2011-25 (Paris: Centre d'Etudes Prospectives et d'Informations Internationales).

Roselló, J., M. Santana-Gallego, and W. Awan (2017). "Infectious Disease Risk and International Tourism Demand," *Health Policy and Planning*, Vol. 32, pp. 538–548.

Santana-Gallego, M., F. Ledesma-Rodríguez, and J. Pérez-Rodríguez (2010). "Exchange Rate Regimes and Tourism," *Tourism Economics*, Vol. 16, pp. 25–43.

Santos Silva, J., and S. Tenreyro (2006). "The Log of Gravity," *Review of Economics and Statistics*, Vol. 88, pp. 641–658.

Tinbergen, J. (1962). "Shaping the World Economy: Suggestions for an International Economic Policy," *The International Executive*, Vol. 5, pp. 27–30.

Wilder-Smith, A. (2006). "The Severe Acute Respiratory Syndrome: Impact on Travel and Tourism," *Travel Medicine and Infectious Diseases*, Vol. 4, pp. 53–60.

Zeng, B., R-W. Carter, and T. De Lacy, 2005, "Short-Term Perturbations and Tourism Effects: The Case of SARS in China," *Current Issues in Tourism*, Vol. 8, pp. 306–322.