Department of Economics and Finance

Gordon S. Lang School of Business and Economics

University of Guelph

Discussion Paper 2022-03

Climate Change and Measures of Economic Growth: Solving the Spatial Mismatch Problem

By:

Devina Lakhtakia
University of Guelph
dlakhtak@uoguelph.ca

Ross McKitrick
University of Guelph
rmckitri@uoguelph.ca

Department of Economics and Finance
Gordon S. Lang School of Business and Economics | University of Guelph
50 Stone Road East | Guelph, Ontario, Canada | N1G 2W1
www.uoguelph.ca/economics

IMPROVE LIFE.

Climate Change and Measures of Economic Growth: Solving the Spatial Mismatch Problem

Devina Lakhtakia*[1] and Ross McKitrick†[1]

[1]Department of Economics and Finance, University of Guelph

*Corresponding author- Postal Address: Department of Economics and Finance, McKinnon Building Room 743, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada; Email: dlakhtak@uoguelph.ca

†Email: rmckitri@uoguelph.ca

Abstract

A challenge in measuring the economic consequences of climate change is that climate is measured at a grid cell level while economic data are provided at the national level. Aggregating climate data up to the national level has been a common approach but the results are sensitive to how the averaging is done and the averaging process itself can bias the results. Here we investigate the climate-economy linkage by downscaling economic data to the grid cell level. Nordhaus (2006) developed the G-Econ database to do this, but while it provides considerable spatial detail it provides only four quinquennial observations per cell from 1990 to 2005. We develop two models to predict within-grid cell economic activity using national, regional and local economic activity. The latter is measured, first, using annual flight volumes at hundreds of urban and rural airports worldwide from 1976 to 2010, and second using satellite measurements of night-time lighting. Nightlights provide more spatial detail but cover a shorter time span than the flights data. Both models exhibit high levels of explanatory power and can be used iteratively to infill and extrapolate the G-Econ data base to the annual level. The flights data, for instance, yields a data set covering grid cells in approximately 150 countries over the 1976 to 2010 interval. Using the newly constructed datasets, we examine the effect of temperature and precipitation increases on income. The relationship changes over time and is generally small and positive, though it depends on the density of production in a grid cell.

Keywords: Gross cell product, climate change, nighttime lights, income

JEL Classification: Q54, Q56, R11

Contents

1 Introduction

Climate change is expected not only to affect natural ecosystems but also economic activities. Studies have found that rising temperatures have lead to a decline in crop yields, labor productivity, a fall in labor working hours (Burke et al. (2015); Graff Zivin and Neidell (2014)) and other economic conditions. Dell et al. (2012) henceforth DJO12, and Zhao et al. (2018) showed that rising temperatures and changes in precipitation patterns can have an effect on not only a country's income but also its economic growth rate. The empirical evidence remains contested however. Newell et al. (2021) show that effects which appear significant disappear when model uncertainty is accounted for. Jiao et al. (2021) present evidence that significant effects may also reflect bias due to the presence of influential outliers in the data.

Studies of the effects of climate change on economic activity combine economic data, which is mainly available at the national level, and climate data, which is mainly available at the local grid cell level (such as one degree by one degree). In order to merge the different data types it is necessary either to break economic data down to the grid cell level or average climate data up to the national level. The latter approach has been more common, but this gives rise to two problems. First, a weighting scheme must be chosen to perform the aggregation, and as noted in Zhao et al. (2018), if the subnational relationship between temperature and economic growth is nonlinear, Jensen's inequality implies that the weighted average will be a biased estimate of the underlying relationship. Second, in practice, results have been found to be quite sensitive to the averaging method. For example the negative effects reported in DJO12 are dependent on using population weights to average gridded temperature data up to the national level. When they use area weights instead the effect of warming on growth becomes smaller and less significant for poor countries, while for the rich countries they either fall to nearly zero or become positive and significant depending on the temperature lag length used.

This paper takes a different approach, reconciling economic and climate data by downscaling output and income to the grid cell level, thereby eliminating the need to choose a national-scale averaging rule. Economic activity corresponding to gross domestic product within a grid cell is called Gross Cell Product (GCP). We make use of the G-Econ data base (Nordhaus, 2006), a $1^\circ \times 1^\circ$ square grid cell-level archive estimating GCP for land-based grid cells worldwide for the years 1990, 1995, 2000 and 2005. The total number of grid cells is 25,572 but the number with non-zero income is below 16,000. We implement an empirical infilling model to predict GCP as a function of national and regional output, geographic and climatic variables, and grid cell-level economic activity. The latter is measured using annual air traffic data from 602 airports worldwide which can be geolocated within specific grid cells. Since we have annual observations for all the explanatory variables from 1976 to 2010 we use the regression equation to predict annual values of the G-Econ data for the unavailable years. We then re-estimate the GCP prediction model using the original G-Econ observations and the infilled values, yielding new set of coefficients. This process is iterated until the coefficients converge. The resulting unbalanced panel consists of 9,349 observations across 150 countries.

For comparison we also form another data set using satellite retrievals of surface nighlight intensity. It is well documented that nighttime lights can be used as a measure of economic activity at the grid cell level. Thus we form a dataset using nighttime lights as the regressor for measuring grid cell activity. This dataset contains 300,000 observations across 158 countries but is only available in the post-1992 interval.

This gives us two comprehensive datasets: one which has a greater time series detail (using flight volumes) and the other which has a wider spatial detail (using nighttime lights). Next, we employ these datasets to study the impact of temperature and precipitation changes on local income. We conduct our analysis for all countries in our sample and evaluate marginal effects of increases in temperature and precipitation on income at different GCP percentiles. We find that the simple bivariate relationship between income and temperature changes over time from a monotonic negative one, similar to that shown in Nordhaus (2006), to an inverted-U shape later in the sample. Estimating a more detailed regression model using grid-cell level fixed effects and controlling for other economic covariates shows that the marginal effects of warming on income are generally small, positive and sensitive to the country's income level. Using the flights-based data we find warming has a small positive effect on income beyond 8°C when evaluated at the sample mean GCP level. At the 25th percentile GCP level the marginal effect is positive but declining with warming and at the 75th percentile GCP level the effect is negative up to 12°C and positive beyond that. Generally similar results emerge using the nightlights-based data except that the income effect in the high GCP case becomes positive at a lower temperature level (about 1.5°C). Using both datasets we find precipitation increases have a positive but declining effect on income in relatively drier regions and a negative impact on the income in relatively wetter regions.

The rest of this paper is organized as follows: Section 2 describes in detail the data sources, Section 3 presents the methodology for constructing the airport-based dataset, Section 4 describes construction of the nightlights-derived dataset, Section 5 studies the impact of climate change on income levels and Section 6 concludes.

2 Data Sources

The G-Econ dataset was developed by Nordhaus (2006) and it measures the economic activity for the entire globe measured at a 1° latitude by 1° longitude resolution. The area of the grid cell and its population are also contained within the dataset. Other reported geophysical variables are elevation, roughness of the Earth's surface, soil types etc. It also incorporates data for a few environmental variables like precipitation and temperature which are measured in monthly averages. The variable of interest for our research here is the "gross cell product or GCP" which measures output with a resolution of 1° latitude by 1° longitude. GCP is conceptually similar to gross domestic product (GDP) and gross regional product. The gross cell product is gross value added in a specific geographic region. So, if we were to add up the gross cell product for all the cells within a particular country, it would give us the GDP of the respective country. The gross cell product is reported in purchasing-power-parity (PPP) adjusted 2005 US Dollars. The entire globe has 64,800 grid cells at the $1^{\circ} \times 1^{\circ}$ resolution, out of which G-Econ database has data on 27,442 grid cells. These data are available for the years 1990, 1995, 2000 and 2005.

The methodology followed by Nordhaus (2006) for constructing the GCP measure was different across countries due to limited availability of data in some places. However, most of the data construction for GCP followed the rule: GCP by grid cell = (population by grid cell) $\times$ (per capita GCP by grid cell). Data at three different levels was collected in order to form the measure of GCP: first at the national level, second at the state level and lastly at the county level. Thus, the method of calculation varied across countries and employed regional estimates on gross product, regional income by industry, regional employment by industry and regional urban and rural population or employment with sector-based data on agricultural and non-agricultural income.

For developed countries like the United States, Canada and Australia gross regional product was made available from the National Statistical Agencies. To calculate the gross cell product at a county level, labor income estimates were used and then a spatial re-scaling was applied to convert this data to a grid cell level basis. For the European Union a similar approach was employed. Eurostat provides regional gross value added data for the European countries based on the political subdivisions. Then, the population density was used to convert the regional data to the grid cell level.

To obtain the GCP for middle-income countries, data on employment was utilized in collaboration with the national data in order to determine the output for regional level. For low-income countries like Nigeria and others for which no data are available, population census data was used to estimate the population at the county level. To this the national output was used to determine the output per capita in the agricultural and non-agricultural sector. The output per capita by region was then obtained by merging the two datasets.

Another concept explained by Nordhaus is that of an "economic desert", which refers to the most unproductive parts of the world where the economic activity is almost zero. The economic deserts of the world are located in the cold regions. There are 6,721 grid cells which have zero economic output and almost all of them lie in Antarctica, Greenland, northern Russia and northern Canada. These could be regarded as outliers as average temperature remains below 0 °C. The distribution of Gross cell product (in per capita terms) across the world in 1° latitude × 1° longitude resolution for 1990 and 2005 can be visualized by the following maps:

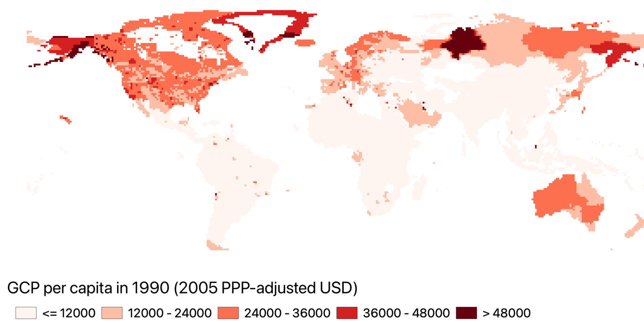

Figure 1: GCP per capita in 1990

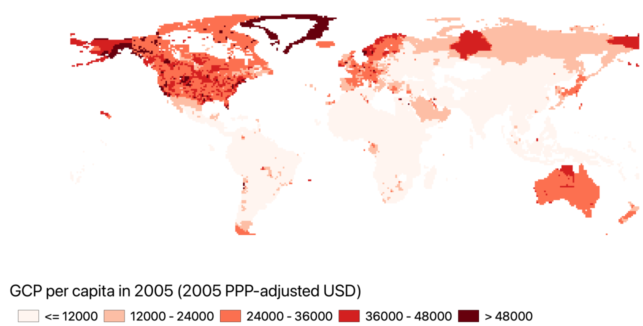

Figure 2: GCP per capita in 2005

We intend to estimate the annual grid cell level output of the G-Econ dataset using as predictors some national and regional aggregates, geographic variables and air traffic data at the grid cell level. The latter choice is motivated by the fact that airline traffic can be precisely located by latitude and longitude in order to match with the G-Econ database. Past studies found a positive relationship between economic growth and airport traffic. Brueckner (2003) shows that a 10 percent increase in passengers using an airline service in a metropolitan area leads approximately to a 1 percent increase in employment in nearby service-related businesses, although not on manufacturing and other goods-related employment. Sheard (2019) reveals that airport size has positive effect on local employment. In a metropolitan area with 1 million residents, a 10% increase in air-traffic leads to 1660 new jobs. This study also concludes like Brueckner (2003) that airport size has large positive effects on employment in the service industry but no significant effect in manufacturing and utilities. So, it seems that airport traffic has an approximate relationship to economic growth which is the reasoning behind taking airline traffic as a proxy for cell level GDP.

The data on economic variables such as national level GDP can be obtained from the Penn World Tables (Feenstra et al. (2015), PWT 8.1). For the data at the regional level, we use the dataset developed by Gennaioli et al. (2014), which covers data on 83 countries and 1,528 regions across the globe and provides the per capita estimates of Regional gross domestic product. The airport traffic data has been sourced from Flight Global for the years 1976-2010, which reports the number of flights originating from different airports but is reported at different frequencies. The population estimates at the grid cell level are available in G-Econ and for the national level can be obtained from PWT 8.1, which can be used to convert the gross national and cell level data on output to a per capita basis. And finally, the data for environmental factors like precipitation and temperature is obtained from Climate Research Unit (CRU)- as Time Series (TS) for the years 1901-2019 (Harris et al. (2020)). It was compiled by The University of East Anglia and it contains data on climate variables across the globe (except Antarctica) measured at 0.5°× 0.5° resolution.

3 Building the Flights-based Data Set

The G-Econ data is available only quinquennially from 1990 to 2005, thus we propose to develop an econometric model to make the data more up to date and to infill the gaps in the existing G-Econ database.

The dependent variable is the cell level Gross cell product in per capita terms (GCPpc) from the G-Econ dataset. We predict it using a regression function of national GDP and regional GDP (in per capita terms), air traffic data and geographic variables (i.e. temperature and precipitation) and at the grid cell level[1].

$$GCPpc_{i,t} = f(NationalGDPpercapita_{i,t}, RegionalGDPpercapita_{i,t}, AirTraffic_{i,t}, GeographicVariables_{i,t}) + e_{i,t} \quad (1)$$

For example, in the year 2000, the GCP per capita for the grid cell at 43° latitude and -79° longitude (city of

[1] We are aware of the potential reverse causality between variables, however this is not a concern for us as our objective is to use them for the sole purpose of prediction

Toronto) would be related to Canada's GDP per capita, Ontario's regional GDP per capita, Toronto Pearson International airport's air traffic, temperature and precipitation (within the airport's grid cell) for the same year. We extract from G-Econ a sample for all the locations across the world for the years 1990, 1995, 2000 and 2005 for which we have airport flight volume data and thereby create our first, quinquennial, dataset. We then run a least-squares regression on the newly created dataset to obtain the coefficients. Using these coefficients, we can then predict the per capita gross cell product for every year for which the independent variables are already available. Then we re-run the regression for all years and use the revised coefficients to re-estimate the explanatory variables. We iterate the infilling and re-estimating process until convergence of coefficients is attained. This gives us our second (annual) dataset.

3.1 The First Dataset

The first dataset contains quinquennial data on Gross cell product along with national levels of GDP, regional level GDP (all in per capita terms), along with data for air traffic and some climate variables for the years 1990, 1995, 2000 and 2005. While G-Econ covers data for 245 countries, due to limited flight data our first dataset covers approximately 150 countries and has 1,160 grid cells in total for the same.

Gross cell product is measured in purchasing-power-parity adjusted 2005 million US Dollars as given by the G-Econ database. The per capita GCP is then calculated by dividing the GCP by the grid cell population. The national GDP obtained from Penn World Tables (PWT 8.1) is measured at purchasing-power-parity adjusted 2005 US Dollars (in millions) and has been converted to a per capita basis by dividing it by the national population. The per capita regional data on GDP is sourced from the Gennaioli et al. (2014) dataset is also measured in PPP-adjusted 2005 US Dollars. The flight data by Flight Global gives the number of flights originating from airports every year. It also gives the latitude and longitude of the airport's location. Each airport has a designated International Civil Aviation Organization (ICAO) and International Air Transport Association (IATA) code for referencing and thus we can easily match which airport lies in which grid cell. A few IATA's were dropped as they ceased to exist or were merged with a bigger airport.

The airports covered in the first datasets are presented in the map below:

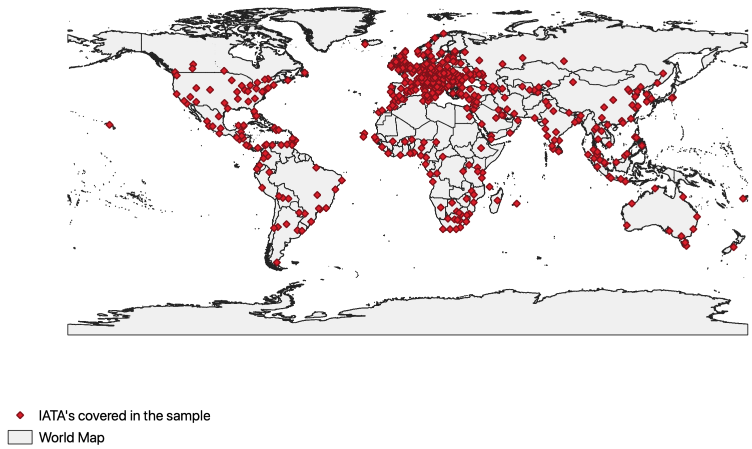

Figure 3: Airports covered in the sample

For our study, we only consider the commercial air traffic and ignore all other traffic movements from the time period 1976-2010. Since the data on air traffic was reported at different frequencies for the various airports, the data was converted to an annual frequency for all the airports. For example, Pula Airport (Croatia) and Boryspil International Airport (Ukraine) report air traffic data on a quarterly basis. N'Djamena International Airport (Chad) and Dushanbe International Airport (Tajakistan) report their data on a semi-annual basis. Airports like Mexico City International Airport (Mexico) and The Rome-Ciampino Airport (Italy) report air traffic data on an annual basis. However, majority of the world's airport like Lester B. Pearson International Airport (Canada), Athens International Airport (Greece), Heathrow Airport (United Kingdom) report monthly air traffic data. Missing data were interpolated in some cases. For annual observations a linear interpolation was applied if only one observation was missing. For data available at monthly frequencies interpolation was used if either one or two months in sequence were missing. Otherwise the missing data were not interpolated.

A sample of the flight traffic data in the year 2005 for six continents can be seen in the maps below:

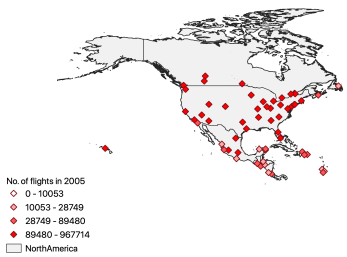

Figure 4: No. of flights in 2005 - North America

Figure 5: No. of flights in 2005 - South America

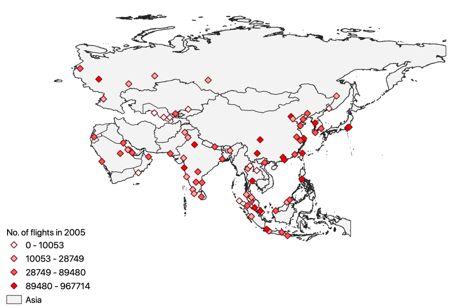

Figure 6: No. of flights in 2005 - Asia

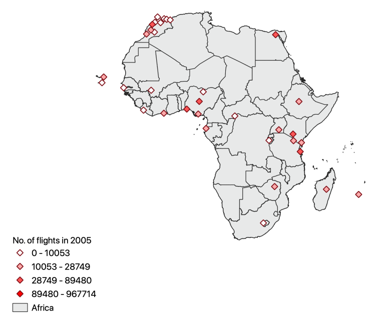

Figure 7: No. of flights in 2005 - Africa

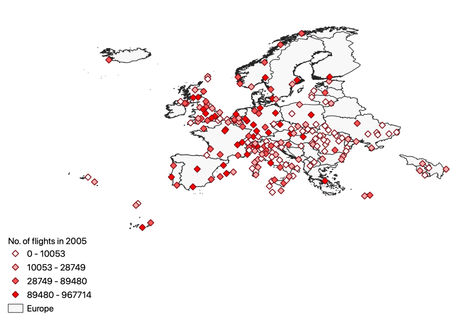

Figure 8: No. of flights in 2005 - Europe

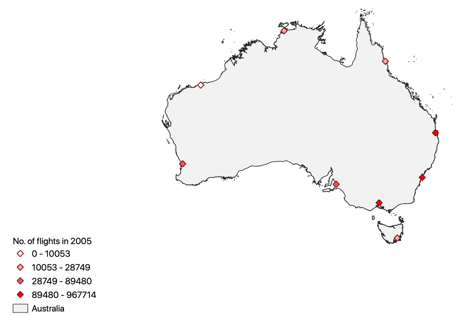

Figure 9: No. of flights in 2005 - Australia

Since the G-Econ data are measured at $1^\circ \times 1^\circ$ grid cells while CRU TS data are measured at $0.5^\circ \times 0.5^\circ$ cells we had to match the spatial scales. We are interested in using mean temperature and precipitation. In the CRU TS data temperature is measured in $^\circ$C and precipitation in millimeters per month. The data for all the variables are available on a monthly basis and were converted to an annual frequency. G-Econ assigns the values of all its variables to the South-West corner of every grid cell. Within each G-Econ cell we have four data points from the CRU TS dataset. All the four data points of CRU TS data within the G-Econ cell were therefore assigned to the South-West corner and then averaged to get one data point for G-Econ. This can be explained via the following diagram:

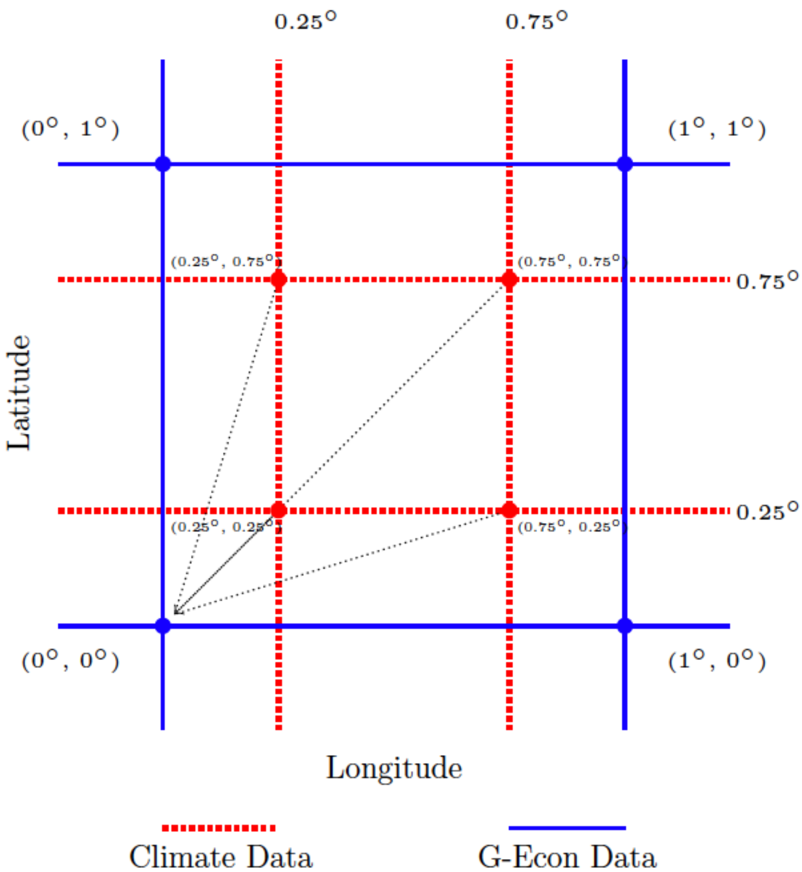

Figure 10: Superimposing climate data on G-Econ data

Suppose we have one grid cell from the G-Econ data (represented by the blue box) measured at 1°× 1°. The CRU TS data is represented by the red boxes which are of 0.5°× 0.5° in resolution. They have an increment of 0.5°, ranging from -179.75° to 179.75° in longitude and -89.75° to 89.75° in latitude.

In Figure 10, we have four data points of the red box represented by (0.25°, 0.25°), (0.25°, 0.75°), (0.75°, 0.75°) and (0.75°, 0.25°) within one G-Econ grid cell. The weather stations report most of the data in anomalies with the exception of precipitation and cloud cover. CRU TS follows an angular distance weighting (ADW) to allocate the the data to the 0.5°× 0.5° grid cells and then this data is converted into absolute terms. For any station to be included in the calculation, it must have atleast 75% of valid observations in a decade. Correlation decay distances (CDD) defines the influence of each station data on a grid cell. Temperature has a long CDD of 1,200 kilometers while precipitation has a CDD of 450 kilometers. Thus, the data in each grid cell for CRU TS data contains data from the number of stations within the grid cell and the data which is interpolated from stations that lie within the CDD of a certain variable. So for our purpose, we only need to consider the four data points in the middle of the G-Econ grid cell as these points contain the station data of the surrounding red boxes had that data have maximum influence on the inner red box. Next, we allocate all the four data points to the South-West corner of the blue box and take the average of those data points. Doing this would give us a single value for G-Econ co-ordinate of (0°, 0°). The rest of the data for other G-Econ grid cells can be obtained in a similar way. The environmental variables utilized in our study are:

1. Mean temperature: The daily average temperature is measured as the average of the day-time high and night-time low values. The yearly value for average temperature was calculated by averaging across all twelve month data and is reported in °C.

   A comparison between the changes in mean temperature between 1976 and 2010 can be visualized via the following graphs:

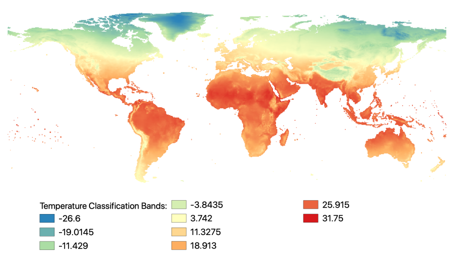

Figure 11: Mean temperatures in 1976

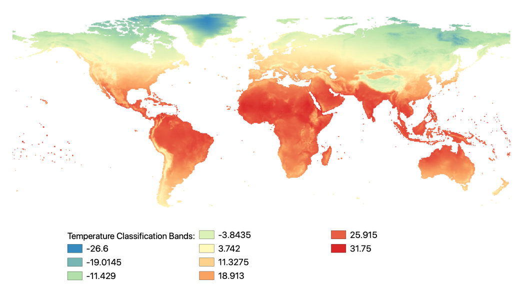

Figure 12: Mean temperatures in 2010

2. Precipitation: Precipitation data is measured in millimeters/month. Thus, to get a yearly value, the precipitation data for all months was added.

The regional GDP per capita obtained from Gennaioli et al. (2014) presents data for different countries for different years, as per the availability of regional data. Since several countries had data only for a few years, we take the regional GDP data for countries which have data available for most years. With this limitation, we interpolate the missing regional GDP data in order to fill up gaps in the dataset for up to a maximum of 5 consecutive missing years of data. For countries which have still have scant data/do not exist in the Gennaioli et al. (2014) dataset or are small countries, we replace the regional GDP per capita with their national GDP (in per capita terms).

The national GDP data in PWT (8.1) is available for the desired time period in PPP-adjusted 2005 million US Dollars. We use a regression model to explain the relationship of Gross cell product per capita (GCPpc) with national level GDP per capita, regional GDP per capita, temperature, precipitation and airport data at the cell level. We found the relationship between GCP and air traffic data is positive and significant in rich countries but not detectable in poor countries so we allowed the model to estimate separate effects for the two groups. We applied the approach used by Dell et al. (2012) to implement this categorization, where we categorize the country as a rich/poor country based on the median income calculated when the country first came into existence in our sample. The variable $Dummy^{RichCountry}$ captures this, taking the value 0 in case of poor country and 1 for a rich country.

The econometric model can be specified as:

$$\begin{aligned} \ln GCPpc_i = \beta_0 + \beta_1 \ln NationalGDPpercapita_i + \beta_2 \ln RegionalGDPpercapita_i + \beta_3 \ln Flight_i \\ + \beta_4 Year_i + \beta_5 Dummy_i^{RichCountry} \times \ln Flight_i + \beta_6 Temperature_i + \beta_7 Precipitation_i + \epsilon_i \end{aligned} \tag{2}$$

All variables are in their natural logarithmic form except for the year and the climate variables. The intercept term takes the form of location (country-level) fixed effects. The independent explanatory variables are: National level GDP per capita (PPP-adjusted 2005 US Dollars), Regional level GDP per capita (PPP-adjusted 2005 US Dollars), air traffic as measured by number of commercial flights taking off from a particular airport at a given time, Year which ranges from 1990-2005, an interaction term between dummy variable for classifying a country as rich/poor times the log of flights, temperature and precipitation. Other variables such as national level of consumption and capital were also considered in the process of building the econometric model, however consumption did not have an explanatory power while capital did not contribute to much variation in GCP when national GDP was included in the regression.

The summary statistics of all the variables involved in the model are presented in Table 1.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
$GCPpc$	17185.06	13965.45	271.63	69582.98	1160
$NationalGDPpc$	16273.93	12926.29	406.35	70252.52	1160
$RegionalGDPpc$	17613.91	15121.5	406.3528	94612.30	1160
$Flight$	74783.55	119971.9	10	967714	1160
$Temperature$	16.36	7.12	1.32	29.65	1160
$Precipitation$	10.20	7.18	0	52.71	1160

Using this dataset, we run a least-squares regression and obtain the following results (omitting the fixed effects coefficients).

Table 2: Regression Results: First Dataset

Variable	Coefficient	Standard Error	t	P $> \|t\|$
$\ln NationalGDPpc$	-0.058	0.083	-0.702	0.483
$\ln RegionalGDPpc$	0.503	0.038	13.255	0.000
$\ln Flight$	0.074	0.021	3.505	0.000
$Year$	0.010	0.002	4.793	0.000
$Dummy^{RichCountry} \times \ln Flight$	-0.013	0.021	-0.610	0.542
$Temperature$	-0.013	0.003	-3.980	0.000
$Precipitation$	0.005	0.002	2.796	0.005
$Constant$	-17.620	3.849	-4.577	0.000
Country FE	Yes			
N	1160			
R^2	0.9703			

Dependent variable: lnGCPpercapita

Notes: Robust standard errors reported in Column (3)

The results from Table 2 suggest that regional level GDP has a highly significant positive effect on the Gross cell product (all in per capita terms). Moreover, our results show that the GCPpc is also positively impacted by the air traffic in that particular cell. Temperature has a negative significant effect while precipitation has a positive significant effect on the gross cell product. In this specification by allowing for locational fixed effects we do not find a significant difference in the effect of flight traffic on Grid cell product between rich and poor countries. The model has an explanatory power of 97.03%. We are aware of the potential endogeneity between gross cell product and the air traffic data. Hence, we also tried a specification taking the Instrumental Variable approach, using a 1-year lag of air traffic as an instrument for air traffic data. However, the results were almost identical apart from a minuscule change in the coefficient of the air traffic variable in our regression.

The coefficients were then tested for out-of-sample predictive power using a K-fold cross-validation routine. The results indicate that these coefficients can predict withheld fractions with a Root mean-squared-error (RMSE) of 0.307 which implies 69.30% accuracy. Therefore we are confident we can use these coefficients to predict the second dataset.

3.2 The Second Dataset

3.2.1 Description

The period for our second dataset is set from 1976-2010 due to the limited availability of air traffic data. The next step was to use the coefficient values obtained in the first regression and the annual values of the explanatory variables to predict the annual values of GCP per capita from 1976-2010. This yielded the initial version of the second dataset. We then repeated the regression step changing the values for GCP per capita to the original values as given by G-Econ for the year 1990, 1995, 2000 and 2005. This gives us new coefficient values. Iterating this procedure, i.e. by using the new coefficients to predict a new dataset was performed until convergence in the coefficients was achieved, yielding the final version of our second dataset. The GCP for the model was calculated by multiplying GCP per capita (in thousands US$) and the grid cell population. In order to extract GCP for all years, we need the grid cell population data for all the respective years as well, for which we used logarithmic interpolation and extrapolation.

The second dataset has 9,349 observations spanning over 150 countries. It covers 69 grid cells from Europe and Central Asia, 17 from North America and 35 cells from Latin America and The Caribbean in the year 1980. Similarly, for the year 2010 the dataset includes 210 grid cells from Europe and Central Asia, 29 from North America and 20 cells from Latin America and The Caribbean. Other regions included in the dataset are: South Asia, Middle East and North Africa, Sub Saharan Africa and East and the Pacific.

3.2.2 Schematic explanation of developing the data base

The dependent variable is Gross cell product in per capita terms (GCPpc) in cell (i) for year (t) and is available for 1990, 1995, 2000 and 2005 in G-Econ. RHS variables are regional/national economic and geographical variables denoted $X_{i,year}$, airport data denoted $A_{i,year}$. The RHS variables are available annually. The available complete records are shown in black font below; incomplete records are in grey.

$$\vdots$$

$$GCPpc_{i,1997} = X_{i,1997}\alpha + A_{i,1997}\beta + (error)$$

$$GCPpc_{i,1998} = X_{i,1998}\alpha + A_{i,1998}\beta + (error)$$
$$GCPpc_{i,1999} = X_{i,1999}\alpha + A_{i,1999}\beta + (error)$$
$$GCPpc_{i,2000} = X_{i,2000}\alpha + A_{i,2000}\beta + (error)$$
$$GCPpc_{i,2001} = X_{i,2001}\alpha + A_{i,2001}\beta + (error)$$
$$GCPpc_{i,2002} = X_{i,2002}\alpha + A_{i,2002}\beta + (error)$$
$$GCPpc_{i,2003} = X_{i,2003}\alpha + A_{i,2003}\beta + (error)$$
$$GCPpc_{i,2004} = X_{i,2004}\alpha + A_{i,2004}\beta + (error)$$
$$GCPpc_{i,2005} = X_{i,2005}\alpha + A_{i,2005}\beta + (error)$$
$$\vdots$$

A regression is run yielding the coefficient set $(\alpha^*, \beta^*)_1$ that predicts $GCPpc_{i,year}$. This model explains 97.03% of the variance in $GCPpc_{i,year}$. We use this regression equation to predict annual values of $GCPpc$. The result is:

$$\vdots$$
$$GCPpc_{i,1997} = X_{i,1997}\alpha + A_{i,1997}\beta + (error)$$
$$GCPpc_{i,1998} = X_{i,1998}\alpha + A_{i,1998}\beta + (error)$$
$$GCPpc_{i,1999} = X_{i,1999}\alpha + A_{i,1999}\beta + (error)$$
$$GCPpc_{i,2000} = X_{i,2000}\alpha + A_{i,2000}\beta + (error)$$
$$GCPpc_{i,2001} = X_{i,2001}\alpha + A_{i,2001}\beta + (error)$$
$$GCPpc_{i,2002} = X_{i,2002}\alpha + A_{i,2002}\beta + (error)$$
$$GCPpc_{i,2003} = X_{i,2003}\alpha + A_{i,2003}\beta + (error)$$
$$GCPpc_{i,2004} = X_{i,2004}\alpha + A_{i,2004}\beta + (error)$$
$$GCPpc_{i,2005} = X_{i,2005}\alpha + A_{i,2005}\beta + (error)$$
$$\vdots$$

The predicted values are indicated in orange. For the 4 years in which G-Econ values are available they are substituted in to replace the predicted values, then the regression is run again:

$$\vdots$$

$$GCPpc_{i,1997} = X_{i,1997}\alpha + A_{i,1997}\beta + (error)$$
$$GCPpc_{i,1998} = X_{i,1998}\alpha + A_{i,1998}\beta + (error)$$
$$GCPpc_{i,1999} = X_{i,1999}\alpha + A_{i,1999}\beta + (error)$$
$$GCPpc_{i,2000} = X_{i,2000}\alpha + A_{i,2000}\beta + (error)$$
$$GCPpc_{i,2001} = X_{i,2001}\alpha + A_{i,2001}\beta + (error)$$
$$GCPpc_{i,2002} = X_{i,2002}\alpha + A_{i,2002}\beta + (error)$$
$$GCPpc_{i,2003} = X_{i,2003}\alpha + A_{i,2003}\beta + (error)$$
$$GCPpc_{i,2004} = X_{i,2004}\alpha + A_{i,2004}\beta + (error)$$
$$GCPpc_{i,2005} = X_{i,2005}\alpha + A_{i,2005}\beta + (error)$$
$$\vdots$$

This yields a new set of coefficients $(\alpha^*, \beta^*)_2$. The second iteration has an R^2 of 99.65% and a squared correlation between predicted and observed G-Econ $GCPpc$ entries of 96.94%. The coefficients are then used to predict $GCPpc$ once again and the process then iterates until the coefficients no longer change. The iteration steps are:

Step	R^2
1	97.03%
2	99.65%
3	99.65%
4	99.65%
$\vdots$	$\vdots$

The process continues for N iterations until convergence in coefficients is attained.

4 The Nightlights-based Dataset

4.1 Satellite Data Coverage

Many studies including that of Henderson et al. (2011, 2012) have established that satellite-measured nighttime lights can be used as a proxy for economic activity. Such data are available from the US National Oceanic and Atmospheric Administration (NOAA) for the years 1992-2013 at 30 arc second grids. With this in mind we form another version of our dataset using nighttime lights data instead of airport data.

Since the nighttime lights data is measured at 30 arc seconds or 0.008333 degree, we used spatial re-scaling in the geospatial software QGIS 3.14 to aggregate the data to a 1×1 degree grid cell in order to match the G-Econ data. In running our first regression, we leave out the regional GDP per capita as a regressor due to the limited availability in the Gennaioli et al. (2014) dataset. Leaving the regressor did not hamper our conclusion as the results are very similar (see Table 3). Thus, using nighttime lights data in the first regression and infilling the values like before, we end up with a dataset of approximately 301,853 observations from the years 1992-2010.

The process for the dataset formed by using nighttime lights is same as that using the airport data. For the first regression we can specify the econometric model as:

$$\begin{aligned} \ln GCPpercapita_i = \beta_0 + \beta_1 \ln NationalGDPpercapita_i + \beta_2 Year_i + \beta_3 Temperature_i \\ + \beta_4 Precipitation_i + \beta_5 Nighttimelights_i + \epsilon_i \end{aligned} \tag{3}$$

A comparison of results between using night lights data and airport data in the first stage regression are presented in Table 3.

Table 3: Comparison of 1st regression results: Airport vs NTL datasets

	Dependent variable: lnGCPpercapita		
	(1)	(2)	(3)
	Airport Data	Airport Data	Nighttime lights Data
$\ln NationalGDPpc$	-0.058	0.365***	0.359***
	(0.083)	(0.072)	(0.018)
$\ln RegionalGDPpc$	0.503***		
	(0.038)		
$\ln Flight$	0.074***	0.132***	
	(0.021)	(0.025)	
$Year$	0.010***	0.008***	0.019***
	(0.002)	(0.002)	(0.001)
$Dummy^{RichCountry} \times \ln Flight$	-0.013	-0.031	
	(0.021)	(0.026)	
$Temperature$	-0.013***	-0.013***	-0.022***
	(0.003)	(0.004)	(0.000)
$Precipitation$	0.005***	0.003	0.000
	(0.002)	(0.002)	(0.001)
$Nighttimelights$			0.021***
			(0.001)
N	1160	1200	45700
R^2	0.970	0.959	0.919
Country FE	Yes	Yes	Yes

Notes: Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As seen in the above table, when regional GDP is dropped as a regressor, the coefficients resulting in the specifications (1) & (2) are similar in magnitude and also of the same sign, except for national GDP which is now positive and significant. The first regression using nighttime light data has 45,700 grid cells and cross validation yields an RMSE of 0.38. This compromise gives us a broader spatial coverage although the time series portion is limited to 1992-2010.

4.2 The Airport Subset of Nighttime Light Dataset

The last dataset we assimilate is a subset of the nighttime light dataset in which there exist values for both airport and nighttime light data. The econometric specification for this model is:

$$\ln GCPpercapita_i = \beta_0 + \beta_1 \ln NationalGDPpercapita_i + \beta_2 \ln RegionalGDPpercapita_i + \beta_3 Year_i + \beta_4 Temperature_i + \beta_5 Precipitation_i + \beta_6 Nighttimelights_i + \epsilon_i \tag{4}$$

The comparison of our original dataset using airport data and the airport subset of nighttime light dataset is presented in Table 4.

Table 4: Comparison of 1st regression results: Airport vs Airport subset of NTL datasets

	Dependent variable: lnGCPpercapita	
	(1)	(2)
	Airport Data	Airport Subset of NTL Data
$\ln NationalGDPpc$	-0.058	-0.097
	(0.083)	(0.104)
$\ln RegionalGDPpc$	0.503***	0.538***
	(0.038)	(0.039)
$\ln Flight$	0.074***	
	(0.021)	
$Year$	0.010***	0.016***
	(0.002)	(0.003)
$Dummy^{RichCountry} \times \ln Flight$	-0.013	
	(0.021)	
$Temperature$	-0.013***	-0.009**
	(0.003)	(0.004)
$Precipitation$	0.005***	0.006***
	(0.002)	
$Nighttimelights$		0.011***
		(0.002)
N	1160	967
R^2	0.970	0.969
Country FE	Yes	Yes

Notes: Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The first regression for this specification contains 967 grid cells and the cross validation result yields an RMSE of 0.35. As we can see in columns (1) & (2), using nighttime light as the dependent variable for grid cell activity gives very similar results if we used the airport data instead. This confirms that airport data is indeed a good predictor for cell level economic activity. Infilling this dataset using the independent variables for the years 1992-2010 gives a dataset of 6,494 observations.

5 Climate change and Local Income

Now we turn to studying the impact of climate change, specifically, changes in temperature and precipitation patterns on income levels across the globe. Note that income is defined as the GCP per capita in the discussions to follow.

Section 5.1 presents the results employing the dataset built on airport flight volume data as the grid-cell level activity predictor while Section 5.2 presents the results using nighttime lights as the proxy for grid-cell level activity.

5.1 Results Using the Flights-based Dataset

Figure 13 shows a simple scatter plot of income and temperature overlaid with a series of 3rd-degree fitted polynomials for the specific years shown. In 1980 the fit yields a monotonic downward slope, but over time a local maximum emerges that moves up to about 10°C. The variations over time indicate, as would be expected, that temperature is not the only causal factor. Comparing the 1980 and 2005 lines, for instance, the slope moderates and even changes sign in the region up to about 13°C.

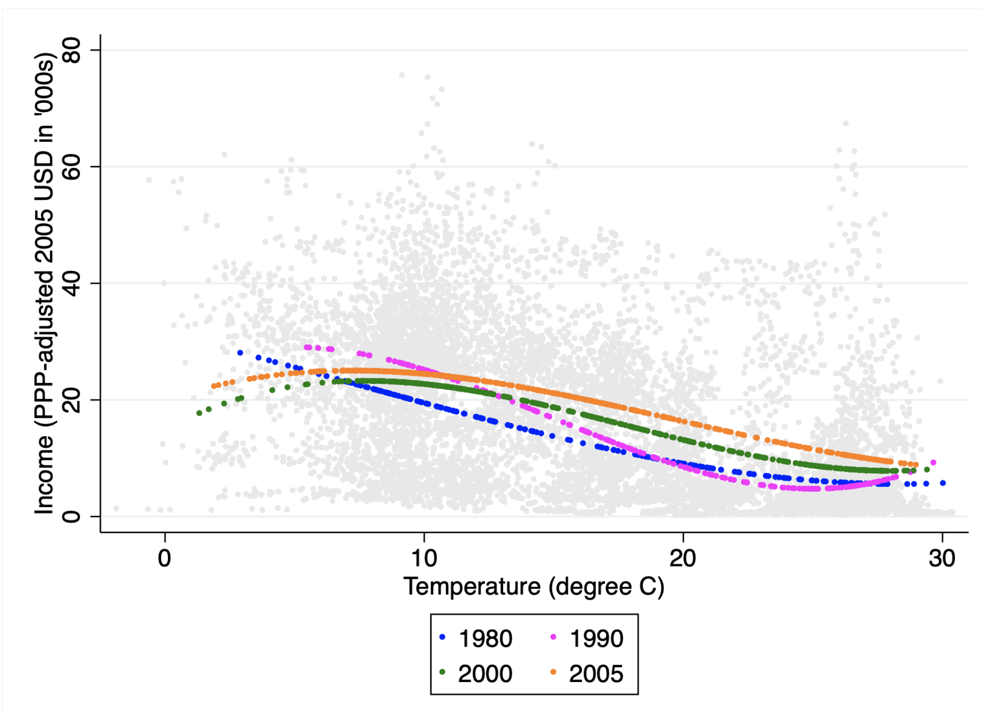

Figure 13: The scatter plot of Income against Temperature: Airport dataset

Our fixed effects panel regression estimating equation includes quadratic terms for temperature and precipitation along with population, lagged grid cell product and interaction terms. Cubic terms were examined but were not statistically significant. The econometric model is specified as follows:

$$
\begin{aligned}
Income_{i,t} = a_i + y_t + +\beta_1 Temperature_{i,t} + \beta_2 Temperature^2_{i,t} + \beta_3 Precipitation_{i,t} \\
+\beta_4 Precipitation^2_{i,t} + \beta_5 Population_{i,t} + \beta_6 GCP_{i,t-1} + \beta_7 GCP_{i,t-1} \times Temperature_{i,t} \\
+\beta_8 GCP_{i,t-1} \times Temperature^2_{i,t} + \beta_9 GCP_{i,t-1} \times Precipitation_{i,t} + \epsilon_{i,t}
\end{aligned} \tag{5}
$$

The fixed effects for location and time are represented by a_i and y_t respectively. Table 5 presents the results for the above model. The fixed effects parameters are not reported but are included in the estimations.

Table 5: Income and Climate Change: Using Airport Data

	Dependent Variable: Income ('000)
$Temperature$	1.4822*** (0.5414)
$Temperature^2$	-0.0172 (0.0153)
$Precipitation$	0.0041*** (0.0010)
$Precipitation^2$	-0.0000** (0.0000)
$Population$	-1.0616*** (0.1692)
$L1.GCP$	0.2544*** (0.0433)
$L1.GCP \times Temperature$	-0.0225*** (0.0057)
$L1.GCP \times Temperature^2$	0.0007*** (0.0002)
$L1.GCP \times Precipitation$	-0.0000*** (0.0000)
N	8396
R^2	0.944
$TimeFE$	Yes
$AirportFE$	Yes

Notes: Robust standard errors in parentheses, adjusted for clustering at grid cell level.
The entries report the results of weighted regression, where the weight is the grid cell population.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

From Table 5, it is evident that coefficients of temperature and the interaction terms of temperature and GCP are mostly significant for all columns. We calculate the marginal effects of temperature and precipitation on income using the derivatives of equation (5) and evaluate them at the sample mean GCP level as well as at the 25^{th} and 75^{th} percentile GCP levels:

$$\frac{dInc}{dT} = \beta_1 + 2\beta_2 MeanTemperature + \beta_7 MeanL1.GCP + 2\beta_8 MeanL1.GCP \times MeanTemperature \quad (6)$$

$$\frac{dInc}{dP} = \beta_3 + 2\beta_4 MeanPrecipitation + \beta_9 MeanL1.GCP \times MeanPrecipitation \quad (7)$$

Note that in this specification the marginal effects are conditioned on more explanatory variables than the simple bi-variate scatter plots in Figure 13, and as a result some different conclusions emerge. The marginal effects of temperature and precipitation using the coefficients from columns (1)-(3) of Table 5 are presented in Figures 14 & 15.

When evaluated at the mean GCP level, after controlling for population and lagged GCP, warming has a positive impact on income beyond 8°C which is significant at temperatures above 10°C. At the 25^{th} percentile of GCP the effect of temperature increases is positive but declining across all temperature range. The effect of warming at the 75^{th} percentile GCP level is negative below 12°C but turns to positive beyond that, with the effects being significant at the lower and higher end temperature range.

Precipitation effects are shown in Figure 15. Note the vertical axis scale is about one one-hundredth that of Figure 14. Increases in precipitation have a positive and significant effect up to 3200 mm of rainfall and a negative effect beyond that when evaluated at the sample mean GCP, with similar results at the lower and higher percentile GCP levels.

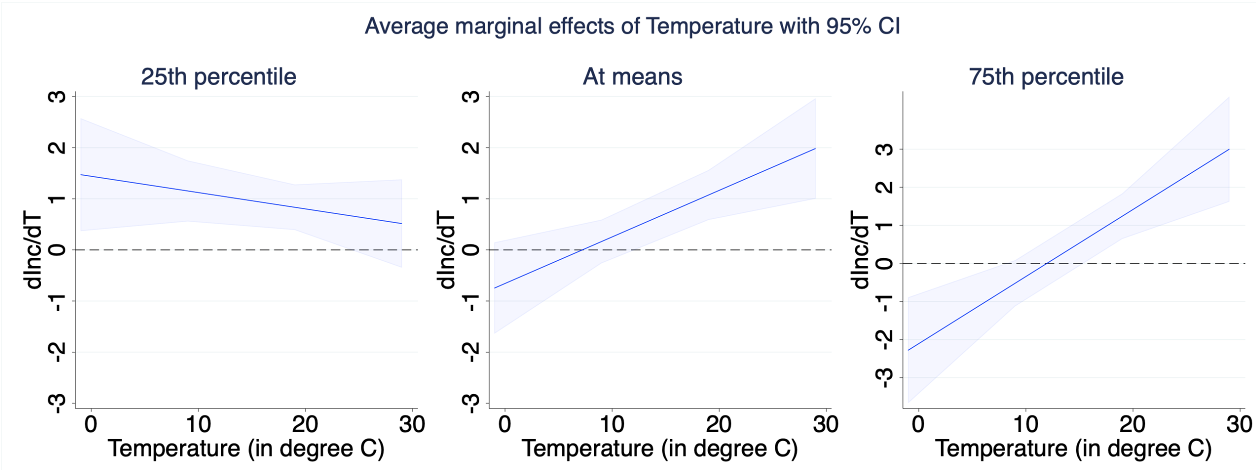

Figure 14: Effect of Temperature on Income Levels (thousands) using Airport Data

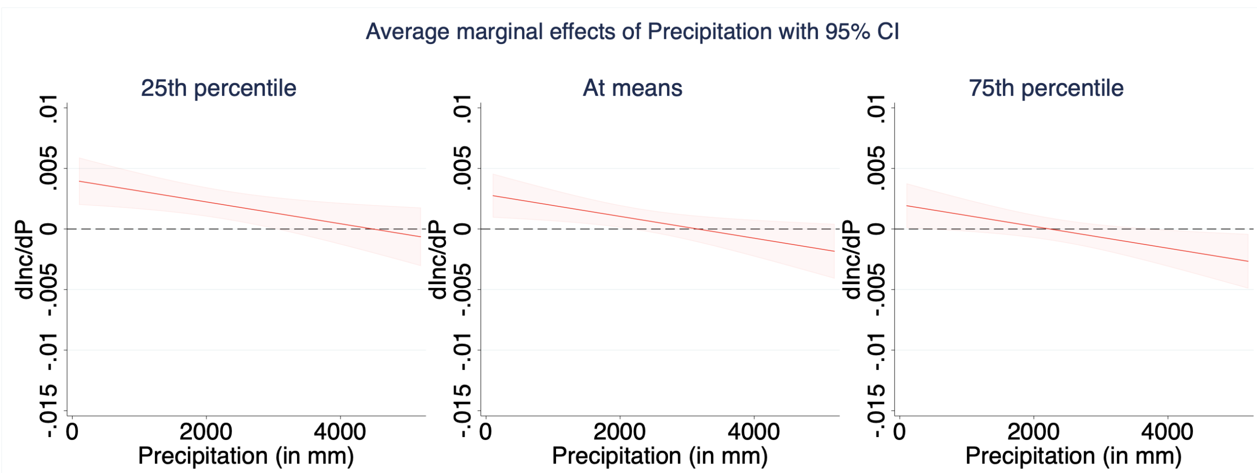

Figure 15: Effect of Precipitation on Income Levels (thousands) using Airport Data

5.2 Results Using the Nightlights-based Dataset

This section analyzes the effect of climate change on Income levels as before but now utilizing the nighttime lights dataset (Section 4.1). Following Section 5.1, Figure 16 depicts a scatter plot of income and temperature and the 3rd-degree fitted polynomials over the years 1992-2010. We truncate the data at 0°C to remove the economic deserts associated with polar and ice-covered areas. Again we see that the income and temperature relationship is not constant over time. The results are concave for all four years but the fitted curve in 1992 is flatter than the one for 2010. This confirms that a single model might not capture the complex relationship between income and temperature.

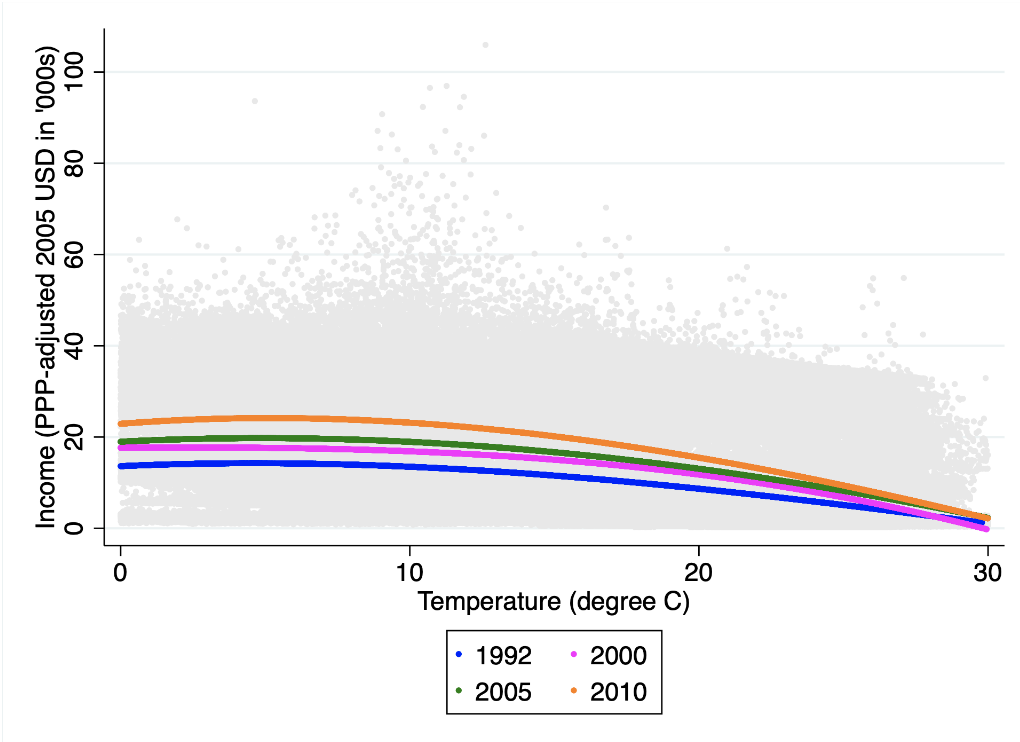

Figure 16: The scatter plot of Income against Temperature: NTL dataset

Table 6 presents the results from running the regression specified in equation 5 and Figures 17 & 18 are constructed based on equations 6 & 7. In order to compare results with the flights-based dataset, we exclude the cold cells from this dataset, namely regions where the mean temperatures are below -2°C. The results using the nightlights data are qualitatively very similar to those from the flights-based data. The effect of warming on income is positive and significant beyond 10°C when evaluated at the sample mean GCP level. When evaluated at the 25^{th} percentile GCP the effect is positive but declining up to 28°C after which it becomes negative, with the effects being largely significant. At the 75^{th} percentile of GCP, the effect is positive beyond 1.5°C and is significant beyond 10°C. Precipitation has a positive and declining effect on income till about 3800 mm of rainfall when evaluated at the 25^{th}, mean and 75^{th} percentile of GCP, consistent with the results from the flight-based dataset.

Table 6: Income and Climate Change: Using Nighttime lights Data

	Dependent Variable: Income ('000)
Temperature	0.9041***
	(0.0931)
Temperature^2	-0.0161***
	(0.0028)
Precipitation	0.0057***
	(0.0005)
Precipitation^2	-0.0000***
	(0.0000)
Population	-0.4482***
	(0.0613)
L1.GCP	0.2587***
	(0.0724)
L1.GCP × *Temperature*	-0.0226**
	(0.0089)
L1.GCP × *Temperature*^{2}	0.0007**
	(0.0003)
L1.GCP × *Precipitation*	-0.0000***
	(0.0000)
N	216547
R^2	0.856
TimeFE	Yes
GridcellFE	Yes

Notes: Robust standard errors in parentheses, adjusted for clustering at grid cell level.
The entries report the results of weighted regression, where the weight is the grid cell population.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

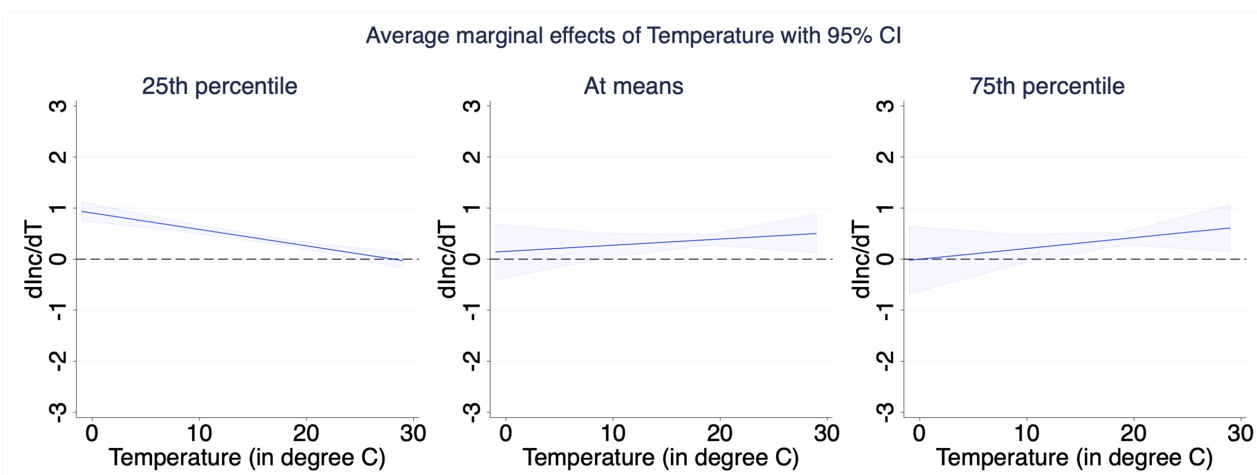

Figure 17: Effect of Temperature on Income Levels (thousands) using NTL Data

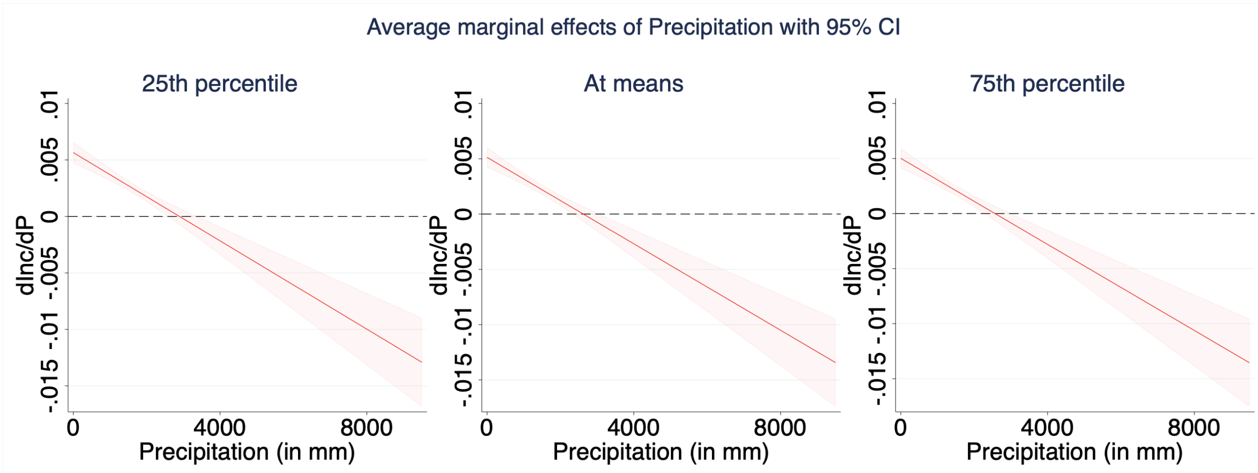

Figure 18: Effect of Precipitation on Income Levels (thousands) using NTL Data

6 Conclusion

Economic and climate data are provided at different spatial scales which can cause difficulty when analyzing the effect of temperatures and precipitation on income. We developed herein an econometric model using national, regional and cell level aggregates which allowed us to extend and infill the quinquennial G-Econ dataset. Using flight volume data for individual airports yields an unbalanced panel of 9,349 annual observations from 1976-2010 across 150 countries. Using satellite-measured nightlights data yields a balanced panel of over 300,000 observations for the 1992-2010 interval. Coherence of the results from the two data sets gives us confidence that they provide a credible empirical basis for estimating the linkages between climate change and economic activity. We found that warming generally has small positive effects on income at the lower percentile GCP level, while at the mean and 75th percentile GCP levels the effect is negative at low temperatures changing to positive above about 10°C. Both data sets also indicate that precipitation increases are associated with higher incomes in the drier half of the data set and lower incomes in the wetter half.

Altogether we find that the relationship between income and climate varies over time and identification depends on controlling for socioeconomic covariates. This points to the importance both of achieving reasonable temporal detail as well as constructing spatially-matched data aggregates that include multiple simultaneous explanatory factors.

References

Brueckner, J. K. (2003). Airline traffic and urban economic development. *Urban studies (Edinburgh, Scotland)*, *40*(8), 1455-1469.

Burke, M., Hsiang, S. M., & Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature (London)*, *527*(7577), 235-239.

Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American economic journal. Macroeconomics*, *4*(3), 66-95.

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the penn world table. *The American economic review*, *105*(10), 3150-3182.

Flight Global. (2012). *Flight global: Pioneering aviation news and insight.* https://www.flightglobal.com/. (Accessed: 2019-09-26)

Gennaioli, N., La Porta, R., Lopez De Silanes, F., & Shleifer, A. (2014). Growth in regions. *Journal of economic growth (Boston, Mass.)*, *19*(3), 259-309.

Graff Zivin, J., & Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of labor economics*, *32*(1), 1-26.

Harris, I., Jones, P., & Osborn, T. (2020). University of east anglia climatic research unit. *https://catalogue.ceda.ac.uk/uuid/89e1e34ec3554dc98594a5732622bce9.*

Henderson, J. V., Storeygard, A., & Weil, D. N. (2011). A bright idea for measuring economic growth. *The American economic review*, *101*(3), 194-199.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *The American economic review*, *102*(2), 994-1028.

Jiao, X., Pretis, F., & Schwarz, M. (2021). Testing for coefficient distortion due to outliers with an application to the economic impacts of climate change. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3915040

Newell, R. G., Prest, B. C., & Sexton, S. E. (2021). The gdp-temperature relationship: Implications for climate change damages. *Journal of environmental economics and management*, *108*, 102445-.

NOAA. (n.d.). *National oceanic and atmospheric administration version 4 dmsp-ols nighttime lights time series.* https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html. (Accessed: 2019-09-26)

Nordhaus, W. D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences - PNAS*, *103*(10), 3510-3517.

Sheard, N. (2019). Airport size and urban growth. *Economica (London)*, *86*(342), 300-335.

Zhao, X., Gerety, M., & Kuminoff, N. V. (2018). Revisiting the temperature-economic growth relationship using global subnational data. *Journal of environmental management*, *223*, 537-544.