The empirics of the tourism-economic growth nexus: How does tourism specialization affect regional growth paths?

David Boto-García[1]
botodavid@uniovi.es

José Francisco Baños-Pino[1]
jbanos@uniovi.es

Roberto Balado-Naves[1]
baladoroberto@uniovi.es

Matías Mayor[2]
mmayorf@uniovi.es

Oviedo Efficiency Group, Department of Economics, University of Oviedo

Department of Applied Economics, University of Oviedo

Abstract:

This paper examines the link between tourism specialization and regional growth rates. Based on macroeconomic growth theory, we use a Bartik instrument to correct for the plausible endogeneity of tourism intensity on GDPpc growth rates. Utilizing data for Spanish provinces in the 2002-2018 period, we find that tourism positively affects regional economic growth. Relevant heterogeneity is nonetheless found depending on the degree of tourism development and specialization in the international segment. We also show that tourism specialization increases employment rates and both the private and public capital stock per capita but negatively affects human capital per worker.

Keywords: tourism specialization; economic growth; tourism-economic growth nexus; Bartik instruments; growth theory

1. INTRODUCTION

Tourism is nowadays a fast-growing service sector in many countries around the world, being an important source of income and job creation (Faber & Gaubert, 2019) and tax revenue (Mohan & Strobl, 2019). After the COVID-19 pandemic, the sector has bounce-back again and tourism arrivals are close to recover their pre-pandemic levels (UNWTO, 2024). It is quite evident that the tourism sector contributes to regional GDP through increased total value added. It is less self-apparent, though, whether the tourism industry is a driving source of long-run economic growth.

A long body of research has provided evidence consistent with the so-called Tourism-Led growth hypothesis, according to which tourism inflows positively contribute to economic growth . However, another stream of literature does not find evidence of a significant relationship. or even argue that tourism specialization puts countries into worser growth paths in what is known as the Beach Disease . As such, whether tourism development contributes to economic growth is yet unconclusive. Moreover, most empirical studies use country-level data (), which masks relevant regional heterogeneity. Research on how tourism specialization affects economic growth at the regional is comparatively scarcer (Bronzini et al., 2022; Cortés-Jiménez, 2008; Lin et al., 2019; Liu et al., 2022; Neuts, 2020; Zuo & Huang, 2018).

The goal of this paper is to formally characterize the relationship between tourism development and subsequent economic growth at the regional level. First, we estimate the contribution of overnight stays per capita to regional growth rates using Bartik instruments to correct for the endogeneity of tourism specialization. Next, we conduct a mechanism analysis to ascertain how tourism specialization affects employment rates, private and public capital deepening and human capital accumulation. In doing so, we look at heterogeneity in the role of regional tourism development on growth rates and input accumulation based on (i) the degree of tourism development of the region (ii) its specialization in the international vs the domestic segment, and (iii) differences between inland and coastal regions.

Our empirical analysis uses data for Spanish provinces in the period 2002-2018. We select Spain as the case study as it stands as the second most visited country in the world and has been subject of great interest in the literature (Balaguer & Cantavella-Jordá, 2002; Inchausti-Sintes, 2015; Liu & Wu, 2019).

The paper has two major contributions to the extant literature on the tourism-economic growth nexus. Firstly, building upon the multisector growth model by Balado-Naves et al. (2024), we estimate the

The second contribution is methodological. Most empirical studies on the link between tourism specialization and economic growth rely on Granger causality tests and time series analyses that in most cases lack theoretical support.[1] The literature that adopts empirical specifications rooted on convergence growth theory typically suffers from endogeneity bias from omitted variables, particularly those using cross-sectional regressions. We, instead, make use of quasi-experimental methods to measure the causal effect of quasi-random variation in regional shares of tourism contribution to the economy on GDP growth using shift-share (Bartik) instruments (Adao et al., 2019; Goldsmith-Pinkham et al., 2020).

The closest work to ours is that by Bronzini et al. (2022). We depart from them in two directions.

2. Theoretical framework

In a recent paper, Balado-Naves et al. (2024) develop a multisector growth model with tourism and non-tourism production to characterize the connection between output growth and tourism specialization. Simplifying the dynamics of the non-balanced growth model of Acemoglu and Guerrieri (2008), with exogenous saving rates of physical and human capital (Solow, 1956; Swan, 1956; Mankiw et al, 1992), the model presents a straightforward way to estimate the net impact of tourism intensity on per capita GDP growth, as well as on the medium- and long-term determinants of economic growth.

Assume that the GDP of a given economic entity (Y) is defined as the aggregate of GDP from touristic activities (Y_{TR}) and that from the remaining sectors (Y_{NTR}) as follows:

$$Y = Y_{TR} + Y_{NTR} \quad (1)$$

[1] One important limitation of these methodologies is that they cannot truly assess the cause-effect relationship between tourism and economic growth; rather, they evaluate the co-movement of variables over time, which in some settings could produce spurious correlations due to omitted variables (Song & Wu, 2022).

If both sectors employ the same type of inputs (physical (K) and human (H) capital stock, and the number of employed people (L)) but face different output elasticities and technologies of production (B_{TR} and B_{NTR}), each sector share on the aggregate pool of primary inputs will be a function of its relative contribution to total output ($sh_{TR} = Y_{TR}/Y$). Accordingly, aggregate GDP can be expressed as a weighted sum of sector technologies embedded in the Neoclassical aggregate production function $F(L, K, H)$:

$$Y = F(AL, K, H) \tag{2}$$

where $A = (sh_{TR}B_{TR} + (1 - sh_{TR})B_{NTR})$ can be read as the aggregate technology of the economy. Since the model presents a stable constant growth path (CGP) in $\left(sh_{TR}^{*}, n^{*}, \tilde{k}^{*}, \tilde{h}^{*}\right)$, where sh_{TR}^{*} is the steady-state value of sector share of tourism over aggregate GDP, n^{*} is the steady-state value of the ratio of employed people ($n = L/P$, where P is total population), and $\tilde{k}^{*}$ and $\tilde{h}^{*}$ are the steady-state values of physical and human capital per effective person $\tilde{k} = K/(AP)$, $\tilde{h} = H/(AP)$; there is a steady-state value of GDP per effective person $\tilde{y}^{*}$.

Following the literature on empirical growth (e.g., Acemoglu, 2009; Barro and Sala-i-Martin, 2004; Islam, 1995), Balado-Naves et al. (2024) show that the growth rate of equation (2) in per capita terms can be linearly approximated around the steady state of the economy as follows:

$$\frac{d \ln y}{d\, t} \approx g_{y}^{*} + \rho_1 \ln sh_{TR-1} + \rho_2(\ln y_{-1} - \ln y^{*}) \tag{3}$$

where g_{y}^{*} is the per capita GDP growth rate in the CGP, and $\ln sh_{TR-1}$ and $\ln y_{-1}$ are the logarithms of the tourism share and per capita GDP in the initial period. Using the fact that $sh_{TR} = Y_{TR}/Y$, and rearranging (3), we have:

$$\frac{d \ln y}{d\, t} \approx g_{y}^{*} + \rho_1 \ln y_{TR-1} + (\rho_2 - \rho_1) \ln y_{-1} - \rho_2 \ln y^{*} \tag{4}$$

Therefore, Equation (4) allows us to test the long-term influence of the volume of tourism services per capita (y_{TR-1}) on per capita GDP growth from a microfounded perspective. Following Acemoglu and Guerrieri (2008), Balado-Naves et al. (2024) state that $\rho_1 > 0$

supports the Tourism-Led Growth hypothesis, because the tourism industry would be more capital intensive and had benefited from relatively faster technological progress. On the contrary, if $\rho_1 < 0$, the Beach Disease hypothesis holds, revealing a weaker tourism sector mostly based on unskilled labor and slower technological progress.

3. Econometric Strategy

Consistent with our theoretical characterization in Equations (1) and (2), we model average annual GDP per capita growth between *t* and *t-s* as a function of a set of covariates at the baseline period, including the ratio of overnight stays to population. Accordingly, the baseline regression equation to be estimated by OLS is:

$$g_{GDPpc_{it}} = \alpha_i + \delta_t + \gamma \ln GDPpc_{it-s} + \beta \ln TOUpc_{it-s} + \theta X_{it-s} + \epsilon_{it} \quad (5)$$

where $g_{GDPpc_{it}} = (\ln \frac{GDP}{POP_{it}} - \ln \frac{GDP}{POP_{it-s}})/s$ is the average annual growth rate of GDP per capita in province/country *i* (for $i = 1, \dots, N$) and period *t* (for $t = 1, \dots, t$) considering an interval of *s+1* years, α_i are province fixed-effects, δ_t are year fixed effects capturing any common temporal shock to all provinces, $\ln GDPpc_{it-s}$ is the log of the GDP per capita *s* years before, $\ln TOUpc_{it-s}$ is the key variable of interest capturing the number of overnight stays per capita, X_{it-s} is a vector of control variables capturing the steady state of the economy, and ϵ_{it} is a random error term.

However, one potential problem of the model in (5) is the likely endogeneity between $g_{GDPpc_{it}}$ and $\ln GDPpc_{it-s}$ on one hand, and $g_{GDPpc_{it}}$ and $\ln TOUpc_{it-s}$ on the other. As discussed in several papers in the applied economic growth literature (e.g., Haulk Jr., 2017; León-González and Montolio, 2015), the level of *GDPpc* at the baseline period likely correlates with the error term through omitted variables. Similarly, β might not be given a proper causal interpretation since tourists might travel to areas with better growth paths (reverse causality). In this respect, some authors have discussed the possibility of a bidirectional causality between tourism development and economic growth. Accordingly, Ordinary Least Squares (hereafter OLS) estimates of equation (5) will deliver inconsistent results. To circumvent this, we move to a quasi-experimental shift-share design.

To get consistent estimates of γ and β, we move to an instrumental variables (IV) framework. One traditional way to proceed is to use lags of the explanatory variables as instruments (e.g., Castelló-Climent, 2010; Llorca-Rodríguez et al., 2021). However, this procedure is highly dependent on the fulfillment of the exclusion restriction; it could even result in more biased estimates than OLS when the lag exerts a direct influence on the outcome variable or on any unobserved confounder captured in the error term (Bellemare et al., 2017; Wang and Bellemare, 2020). Alternatively, we move to Bartik-type shift-share design (Adao et al., 2019; Goldsmith-Pinkham et al., 2020), which is nowadays considered a credible strategy to get causal estimates when working with observational data.

4. Data

We construct a balanced panel dataset of Spanish provinces (N=50, NUTS 3) at the year level for the period 2002-2018 (T=17) involving a total of 850 observations. We collect information at the province-year level on the following variables: (i) total GDP and the corresponding GDP values for agriculture (*GDP_agr*), industry (*GDP_ind*), construction (*GDP_cons*) and services (*GPD_serv*) sectors (all expressed in thousand euros), (ii) number of domestic (*TOU_DOM*), foreign (*TOU_FOR*) and total (*TOU*) tourists, and number of domestic (*STAYS_DOM*), foreign (*STAYS_FOR*) and total (*STAYS*) overnight stays (all in thousand people), (iii) population (*POP*, in thousands), (iv) number of employed and active people (*L* and *N*, in thousands), (v) private and public capital stock on material assets without real state (*Kpriv* and *Kpub*, in thousand euros), and (vi) number and share of employed people with university studies as a measure of human capital (denoted by H and $h=H/L$, respectively). All the monetary variables are expressed in real terms after deflating them by the Consumer Price Index (base 2021).

We define our indicator of tourism intensity as the ratio of overnight stays to population.[2] We also use tourist numbers instead of stays as the numerator as a robustness check. Furthermore, due to data availability reasons, we use data on overnight stays and tourist numbers accommodated in hotel establishments. Unfortunately, information for other accommodations like tourist apartments or rural houses are only available since 2010, which leaves us with a reduced temporal span for the analysis.

[2] Ideally, we would like to have information on the share tourism represents over regional GDP per year. Unfortunately, this information is not available at the province level.