Spatial and Statistical Heterogeneities in Population Science Using Geographically Weighted Quantile Regression

Vivian Yi-Ju Chen[*]   Tse-Chuan Yang[**]

* Associate Professor, Department of Statistics, National Chengchi University. Corresponding author.
  E-mail: viyjchen@nccu.edu.tw
** Associate Professor, Department of Sociology, University at Albany, State University of New York.
  E-mail: tyang3@albany.edu

Abstract

There is a growing interest in exploring heterogeneous associations with independent variables across the distribution of either the dependent variable (using quantile regression) or across geographic space (using geographically weighted regression). The former is often known as statistical heterogeneity, whereas the latter refers to spatial heterogeneity. However, population research has been slow to adopt either of these methods. This study first briefly discusses why more attention to the concept of heterogeneity is needed and then introduces a method that simultaneously considers statistical and spatial heterogeneity, namely geographically weighted quantile regression (GWQR). We illustrate how to use GWQR with U.S. county-level coronavirus disease (COVID-19) vaccination data and explain how GWQR identifies significant heterogeneities in the relationships between the vaccination rate and its determinants across space and over the vaccination distribution. The results suggest that both spatial and statistical heterogeneity are a common occurrence. For example, the COVID-19 case rate has a stronger association in counties in the lower quantiles than in the higher quantiles. The spatial distribution of this relationship is focused on counties in the Mountain states and is shifted to the Midwest region. As such, we conclude that both heterogeneities should be considered in population research.

Keywords: heterogeneity, homogeneity, geographically weighted regression,
quantile regression, spatial demography

I. Introduction

Most population studies seek to identify the associations between a dependent variable and a set of independent variables. The ordinary regression modeling technique is arguably the most common approach (Hao and Naiman 2007) to depict these associations. This approach typically uses the conditional-mean function to summarize how a dependent variable responds to the changes of independent variables (Kleinbaum et al. 2008). While the mean regression can provide a parsimonious picture of the associations of interest, crucial assumptions are required to fit the conditional-mean model, such as that the errors need to be normally distributed with homoscedastic variance and that values of dependent variable must be independent. These assumptions, however, have been often violated in many applications and it is particularly so when analyzing social science and/or geospatial data. (Cressie 1993; Hao and Naiman 2007). Demographers have begun to explicitly address the analytical issues with alternative modeling techniques. For instance, models allowing for heteroscedastic errors then have been considered through robust estimations of error covariance matrix or by means of weighted least squares. The independence assumption is often suspected for data aggregated based on a set of geographical regions over the study area and hence spatial modeling approaches have thrived (Schmertmann et al. 2008, 2010; Voss et al. 2006). Despite these efforts in empirical data analysis, we suggest in this study that one way to move the field of demography forward is to simultaneously explore, within one analytical framework, two sources of heterogeneity in data—the statistical heterogeneity across the distribution of a dependent variable and the spatial heterogeneity with respect to data relationships

concerning instability of regression coefficients in a model (Fotheringham 1997; Harris 2019).

To better understand the concept of heterogeneity, it is imperative to first introduce its opposite, homogeneity. In statistics, homogeneity refers to the level of similarity in distributions and the definition of similarity ranges from a single attribute (e.g., variance or mean) to total sameness. Extending the concept of similarity to a spatial context, the relationships of interest that do not vary with the spatial dimension can be described as homogeneous. For simplicity and convenience, homogeneity is the underlying assumption in most conventional analytic techniques and heterogeneity per se is often ignored or regarded as an unwelcome complication (Kleinbaum et al. 2008; Pickett and Cadenasso 1995).

It follows that heterogeneity can be defined as the lack of homogeneity. A type of statistical heterogeneity is said to exist when the relationships obtained with the conditional-mean model cannot be generalized to other locations along the response distribution, especially for the data skewed or distributed in the tails (Hao and Naiman 2007). The development of quantile regression (QR) attempts to address this limitation (Koenker 2005; Koenker and Bassett 1978). In contrast to traditional regression approaches, QR is flexible with non-normally distributed errors, provides strong robustness against outliers, and more importantly, offers the ability to detect statistical heterogeneity examining how predictors are associated with a dependent variable along its distribution (Abrevaya and Dahl 2008; Austin et al. 2005; Han et al. 2011). In certain circumstances, the mean or median of the response variable is not of interest. Few researchers have taken advantage of QR to explore heterogeneity in demography research. For example, a debate exists whether or not income inequality is a determinant of mortality. Yang et al. (2012) suggested to better clarify the question by identifying

the inequality–mortality relationships with QR than traditional analytic approaches, especially on lower and upper levels of mortality. Similarly, to examine if there is a significant difference in the relationship between economic expansion and migration across labor markets, Partridge et al. (2012) employed QR to achieve the goal.

On the other hand, the essence of demography is to study the collective behaviors and spatial demography focuses on the analysis of demographic processes across space (Weeks 2004). Population processes and demographic changes have been traditionally studied using data aggregated to various geographic units (e.g., counties and states), which is often labeled as the macro-demographer's perspective. Since the 1990s, the development of user-friendly spatial analysis tools has not only reinvigorated spatial demography, but also served to motivate spatial perspectives in demographic policy-making (Voss 2007). A recent report (Matthews et al. 2021) identified several major challenges for spatial demography, including the need for new methodological development in spatial analysis and the integration of these new spatial methods into population research. Spatial econometrics models have been widely employed by demographers but the methods assume that the associations between the independent and dependent variables are homogeneous across space (i.e., do not vary by location), thus overlooking the so-called spatial heterogeneity with respect to data relationships (Fotheringham 1997; Harris 2019). Geographically weighted regression (GWR) is a method that can facilitate such understanding of spatial heterogeneity. GWR provides researchers a more thorough picture of underlying spatial process between independent and dependent variables and, in turn, helps explore place-specific processes (Brunsdon et al. 1998; Fotheringham et al. 2002; Wheeler and Páez, 2010). Some examples in population studies include Partridge and Rickman (2007) which used GWR

to identify persistent-poverty areas and suggested that place-based policy could play a crucial role in reducing poverty in those areas, and Gutiérrez-Posada et al. (2017) which applied GWR for modeling of local population growth.

Compared to other disciplines such as forestry, geography, economics, and environmental science, demography has lagged in exploring the heterogeneities with the methods QR and GWR discussed above. While increasing efforts have been made to adopt either QR or GWR into population research, we argue that it remains necessary to explore at the same time both statistical and spatial heterogeneity. Scholars have developed a method that claims to achieve the task in this endeavor, namely geographically weighted quantile regression (GWQR) (V. Y.-J. Chen et al. 2012). Building upon this new methodological development, this study aims to discuss why population researchers should care about heterogeneity and illustrate how to use GWQR in population studies. Below, we concisely discuss QR and GWR and then explain how GWQR synergizes both methods. One case study with U.S. county-level coronavirus disease (COVID-19) vaccination data will be used as an illustration, followed by the discussion of the future of GWQR in population studies.

II. Literature Review

QR, GWR, and Some Limitations of Traditional Regression Approaches

Linear regression is commonly used to answer the question of how a continuous dependent variable is associated with a set of independent variables in population studies (Kleinbaum et al. 2008). There are, however,

several well-known disadvantages of this approach. The normality assumption (i.e., data follow the Gaussian distribution) may not hold especially when a dependent variable has a heavy-tailed distribution (Hao and Naiman 2007). The phenomena of dependence (i.e., data are not randomly distributed) is often occurred in ecologic studies (Cressie 1993). Failing to meet these modeling assumptions may result in biased estimates, incorrect inferences, and misleading conclusions. To address these issues, new statistical methods have been developed.

Beyond these assumptions, the focus on modeling mean value may have inadvertently steered demographers away from questions relevant to non-central locations (Hao and Naiman 2007). For example, as suggested by Machado and Mata (2005), understanding the relationship between wage inequality and educational attainment in the tails of the wage distribution may be more helpful than models estimated for the central locations. Likewise, the associations of maternal characteristics with birth weights should be more important in the low end (among low birth weight infants) than any other locations of the birth weight distribution (Abrevaya and Dahl 2008). The traditional approach inherently fails to characterize the relationship between a dependent variable's distribution and independent variables, and the question of how changes in independent variables affect the shape of a dependent variable's distribution is unanswered (Koenker 2005).

By contrast, QR has drawn much attention because it focuses on estimating the relationships at any location of the distribution of the dependent variable (Koenker 2000). As such, QR has emerged in health and population research (Austin et al. 2005; Yang et al. 2012). QR divides the whole distribution into quantiles and estimates the conditional quantiles as functions of explanatory variables. Quantile is a term that generalizes

specific locations, such as quartile, quintile, decile, and percentile, but it can represent any predetermined locations of a distribution (Koenker and Bassett 1978). Since any quantiles can be modeled, researchers can choose any positions in a response distribution to tailor specific research questions, therefore portraying varying impacts of independent variables on the dependent variable as a form of statistical heterogeneity (Koenker 2005).

Demographers have realized that spatial dependency has commonly embedded in ecological data (Tobler 1970) and such dependency violates the independent assumption. With the resurgence of spatial demography (Voss 2007), spatial econometric has been widely adopted to address spatial dependency (Chi and Zhu 2008). In general, spatial dependency is accounted for with a spatial lag effect of the dependent variable and/or a spatial error term in the residuals. This approach incorporates the adjacency of any two spatial units in the construction of a spatial weight matrix (LeSage and Pace 2009). However, spatial econometric methods mainly estimate a single, or global, regression equation in which the relationships between an independent and the dependent variable are assumed to be stationary over space (Fotheringham 1997). Such model specification does not allow for the exploration of spatially varying relationships, a core property of spatial data termed as spatial heterogeneity with respect to data relationships (Harris 2019). Some researchers have applied geostatistical models (e.g., semivariograms) to ecological data and adjusted for spatial dependence based on the distance between two units (Schmertmann et al. 2008), but most of the techniques are essentially exploratory and do not include explanatory covariates into analysis.

GWR has been proposed to explore spatial heterogeneity of regression relationships between the dependent and independent variables. Because of its ease of computations, scholars have advocated the use of GWR to

understand the spatially varying association despite some concerns, such as collinearity among local estimates (Wheeler and Páez 2010; Wheeler and Tiefelsdorf 2005), and it has found considerable applications to population and health studies in the past decade (e.g., V. Y.-J. Chen et al. 2010, 2022; Gebreab and Diez Roux 2012; Mollalo et al. 2020; Nakaya et al. 2005; Shoff et al. 2012, 2014). Not only does GWR account, in part, for spatial dependence, but also helps identify potential spatial heterogeneity (place-specific relationships) underlying the ecological data (Fotheringham et al. 2002).

It should be noted that there exist various definitions of heterogeneity, and each of the definitions relates to the object to which heterogeneity applies (Dutilleul and Legendre 1993). In the traditional regression modeling, for example, heteroscedasticity refers to the equality of variances (Dutilleul and Legendre 1993; Glaser 1983; Sokal and Rohlf 1981) and this has been found to be one consequence of heterogeneity. Hierarchical modeling that considers the between-unit heterogeneity (and within-unit dependencies) with random effects (Raudenbush and Bryk 2002) may present a way to resolve the issue in such case. Different from such consideration, the present study draws emphases in estimating the varying relationships at any location of the distribution of the dependent variable and across space, referring to the statistical and spatial heterogeneity indicated previously. The subsequent section will discuss why a method that accounts for both statistical and spatial heterogeneity is important.

Why is GWQR Desirable?

Several other spatial methods have been developed to handle spatial heterogeneity in regression relationships, such as spatial regime modeling (Anselin 1988a, 1988b) and spatially varying coefficient modeling (Gelfand

et al. 2003). The former is essentially a full interaction modeling approach that examines whether there is any structural difference across regimes, which are pre-defined by researchers (Anselin 1988a, 1988b) and generates parameter estimates for each regime. Compared with this approach, GWR heavily focuses on local spatial processes (Waller et al. 2007), does not require pre-defined regimes/areas, and can be used to interpolate local estimates at locations without data (Fotheringham et al. 2002; Gelfand et al. 2003). Consequently, GWR is more "geographically" flexible and can provide information on more detailed place-specific relationships (spatial heterogeneity with respect to data relationships) than spatial regime models can. Spatially varying coefficient modeling represents as a parallel of GWR, where the spatially varying regression parameters are accommodated through a multivariate spatial process within Bayesian hierarchical modeling framework. Though possibly processing better inferential properties, the methodology relies on the specification of prior distributions and is sometimes computational demanding. It also has not been widely adopted by population researchers as opposed to GWR. Detailed discussions regarding the comparisons between GWR and spatially varying coefficient models can be found elsewhere (Wheeler and Calder 2007; Wheeler and Waller 2009; Finley 2011).

While GWR is a popular technique for exploring spatial heterogeneity with respect to data relationship, its current structure only computes the parameter estimates for the relationships of the mean value of dependent variable with the independent covariates and fails to consider other locations in a dependent variable's distribution. This shortcoming is inherent in GWR approach as it is expanded from the conventional ordinary least squares regression.

To answer more complex research questions, it would be necessary to

develop a method that allows one to obtain the relationships of independent variables with a particular quantile of interest across space. That is, there are likely many advantages to the integration of QR in a spatial analysis framework. Several leading spatial statisticians have noticed this need and developed analytical tools towards spatial quantile-based regression analysis. Examples include the Bayesian spatial QR models extended based on spatially varying coefficient modeling (Reich et al. 2011) or with the asymmetric Laplace process to incorporate spatial dependence in the variance–covariance structures (Lum and Gelfand 2012). Such methods using Bayesian perspective need to be implemented with Markov chain Monte Carlo algorithms, which may be a barrier to applying these spatial QR methods to empirical demographic studies. In light of this issue, V. Y.-J. Chen et al. (2012) proposed GWQR by integrating the features of QR into GWR and handle the spatial associations among observations with a kernel density function. GWQR is based on a nonparametric perspective and could be implemented with SAS® (V. Y.-J. Chen and Yang 2012) or R.

Building upon the strengths of QR, GWQR allows users to examine how certain independent variables affect the dependent variables across the distribution of the dependent variable (i.e. statistical heterogeneity). And this location-specific relationship can be extended to space (i.e. spatial heterogeneity). How do these heterogeneities learning from GWQR inform demographic and health research? At its most basic, and following Fotheringham et al. (2002), GWQR challenges the conventional assumption that the same stimulus promotes the same response. There are several reasons that bolster this proposition. First, heterogeneity should be regarded as an inherent feature of a society (Dutilleul and Legendre 1993). Many social characteristics are unevenly distributed and people are embedded in different geographic scales that affect spatial heterogeneity (Matthews

2011). Indeed, it is naïve to assume that the relationships of interest are constant everywhere (Lloyd 2010). Second, heterogeneity may reflect the population dynamics with the natural and/or social environment that have not been considered by researchers. This unobserved association may result in heterogeneous relationships across regions and encourage scientists to adopt a local perspective and locate areas that need special attention (Anselin 1988a; Dutilleul and Legendre 1993). Moreover, the spatial pattern generated by heterogeneity provides valuable information to those researchers attempting to identify factors that are predictive of the spatial pattern (Fotheringham et al. 2002; Lloyd 2010). Finally, there is a need in many instances for new methods that allows demographers to answer questions unable to be addressed by QR and GWR in isolation. Taking body mass index (BMI) as example, obesity researchers are particularly interested in exploring the factors for high, rather than the average, BMI (C.-M. Chen et al. 2012). Coupled with the fact that it is argued that space plays an important role in the prevalence of obesity (Drewnowski et al. 2007), the investigations of how the dependent variable reacts to the change in an independent variable at quantiles of interest (other than mean) across space may be then helpful, providing new insights into heterogeneity.

In sum, GWQR combines QR and GWR, which makes it possible to simultaneously account for the heterogeneities across space and the distribution of a dependent variable (V. Y.-J. Chen et al. 2012). The novel analytic approach allows researchers to dissect the global processes into local processes and investigate whether and how the relationships between predictors and the dependent variable differ across the distribution of the dependent variable. The original paper (V. Y.-J. Chen et al. 2012) describes the technical details but a brief discussion of GWQR follows.

III. Methodology

GWQR Framework

To understand GWQR, a basic understanding of Gaussian GWR modeling is necessary. Following Fotheringham et al. (2002), a Gaussian GWR model can be expressed as

$$Y_i = X_i^t \beta(u_i, v_i) + \varepsilon_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} X_{ik}\beta_k(u_i, v_i) + \varepsilon_i \quad \text{........................(1)}$$

where Y_i, $i = 1, 2, \ldots, n$, represents the response observations collected from location i in space and the corresponding geospatial covariate vector can be written as $X_i = [1, X_{i1}, \ldots, X_{ip}]^t$ of dimension $(p + 1)$ including the constant 1 for intercept. $\beta(u_i, v_i) = [\beta_0(u_i, v_i), \beta_1(u_i, v_i), \ldots, \beta_p(u_i, v_i)]$ indicates the local coefficient estimates, (u_i, v_i) represents the coordinates of location i in space, and ε_i is the error term with a mean of zero and common variance σ^2. The coefficient estimations, β_k, are specific to each location i, which yields mappable local statistics.

One feature of GWR is the ability to use a kernel-based density function centered on each observation to estimate the local parameters. That is, the β's at a given location (u_i, v_i) are locally estimated through the following weighted least square function

$$\sum_{j=1}^{n} \left[Y_j - \beta_0(u_i, v_i) - \sum_{k=1}^{p} X_{jk}\beta_k(u_i, v_i) \right]^2 w_{ij} \quad \text{.......................................(2)}$$

where $w_{ij} = K(d_{ij}, h)$ is the geographical weight assigned to each observation and calculated in terms of a kernel function K, depending on the distance d_{ij} between the given location (u_i, v_i) and the jth designated location $(u_j,$

v_j). The kernel size is also regulated by a bandwidth h, which controls the smoothness of the estimates and can be either fixed or adaptive. Overall, more weight is placed on observations closer to (u_i, v_i) than those farther away. Several choices exist for the kernels; see Fotheringham et al. (2002) for more descriptions. Typically, the adaptive bisquare for h being determined by a nearest-neighbor approach is frequently used in literature:

$$w_{ij} = \begin{cases} \left[1 - \left(\dfrac{d_{ij}}{h} \right)^2 \right] & if\ d_{ij} < h \\ 0 & otherwise \end{cases} \quad \text{..(3)}$$

Fotheringham et al. recommend such weighting scheme as it avoids calibration problems especially when the data are sparse or irregularly distributed over space.

GWQR extends Equation (1) based on the framework of QR, defined as

$$Y_i = X_i' \boldsymbol{\beta}_\tau(u_i, v_i) + \varepsilon_{\tau i} = \beta_{\tau 0}(u_i, v_i) + \sum_{k=1}^p X_{ik}\beta_{\tau k}(u_i, v_i) + \varepsilon_{\tau i} \quad \text{...................(4)}$$

where the vector $\boldsymbol{\beta}_\tau(u_i, v_i) = [\beta_{\tau 0}(u_i, v_i), \beta_{\tau 1}(u_i, v_i), \ldots, \beta_{\tau p}(u_i, v_i)]^t$ now represents the coefficients at different quantiles ($0 < \tau < 1$) for each location (u_i, v_i). $\varepsilon_{\tau i}$ denotes the random error whose τ^{th} quantile conditional on X_i is equal to zero. In contrast to the original GWR Equation (1), the parameters of GWQR, $\boldsymbol{\beta}_\tau(u_i, v_i)$ depend on location coordinates and also the quantiles τ. According to V. Y.-J. Chen et al. (2012), these coefficients are estimated nonparametrically using kernel smoothing methods similar to that of GWR. Specifically, the local estimators can be obtained by minimizing

$$\sum_{j=1}^{n} \rho_\tau \left[Y_j - \beta_{\tau 0}(u_i, v_i) - \sum_{k=1}^{p} X_{jk}\, \beta_{\tau k}(u_i, v_i) \right] w_{ij} \quad \text{.................................(5)}$$

where $\rho_\tau(z) = z(\tau - I(z < 0))$ is known as the "check" loss function with $I(\cdot)$ being the indicator function, and w_{ij} is defined as previously.

V. Y.-J. Chen et al. (2012) consider the solutions of Equation (5) as local constant estimators based on the theory of local modeling. Alternatively, one may have the local linear approximation in a neighborhood of given location (u_i, v_i) with Taylor's expansion by

$$\beta_{\tau k}(u, v) \approx \beta_{\tau k}(u_i, v_i) + \beta_{\tau k}^{(u)}(u_i, v_i)(u - u_i) + \beta_{\tau k}^{(v)}(u_i, v_i)(v - v_i) \quad \text{..................(6)}$$

in which $\beta_{\tau k}^{(u)}(u_i, v_i)$ and $\beta_{\tau k}^{(v)}(u_i, v_i)$ are the continuous partial derivatives of $\beta_{\tau k}(u, v)$ at the location (u_i, v_i) with respect to the coordinates u and v (N. Wang et al. 2008; V. Y.-J. Chen et al. 2012); $k = 0, 1, \ldots, p$. The GWQR calibrated using the approximate expression of Equation (6) is considered as local linear GWQR. We let the coefficient vectors $\boldsymbol{\beta}_\tau^{(u)}(u_i, v_i) = [\beta_{\tau 0}^{(u)}(u_i, v_i), \ldots, \beta_{\tau p}^{(u)}(u_i, v_i)]^t$, $\boldsymbol{\beta}_\tau^{(v)}(u_i, v_i) = [\beta_{\tau 0}^{(v)}(u_i, v_i), \ldots, \beta_{\tau p}^{(v)}(u_i, v_i)]^t$, and $\boldsymbol{\beta}_\tau(u_i, v_i) = [\beta_{\tau 0}(u_i, v_i), \ldots, \beta_{\tau p}(u_i, v_i)]^t$. Based on Taylor's theorem, the local linear GWQR estimators for $\boldsymbol{\beta}_\tau(u_i, v_i)$ are the corresponding solutions that minimize the following loss function:

$$\sum_{j=1}^{n} \rho_\tau \left\{ Y_j - X_j^t \left[\boldsymbol{\beta}_\tau(u_i, v_i) + \boldsymbol{\beta}_\tau^{(u)}(u_i, v_i)(u_j - u_i) + \boldsymbol{\beta}_\tau^{(v)}(u_i, v_i)(v_j - v_i) \right] \right\} w_{ij}$$
$$\text{...(7)}$$

As shown above, the difference between local constant (Equation [5]) and local linear (Equation [7]) models lies in how they estimate the local relationships in the neighborhood of observations. The local constant approach is relatively simple to implement, whereas the local linear estimator is more flexible and allows the coefficients to be expressed as

linear functions of the spatial coordinates. Such linear approximation may reduce bias and result in a better fit compared to the local constant estimator at the cost of computational efficiency (N. Wang et al. 2008; Páez et al. 2011).

It is known from the theory of QR that the weighted QR problem of Equations (5) or (7) can be equivalently reformulated as a problem of optimization in terms of linear programming (C. Chen and Wei 2005; Koenker 2005). The estimates $\hat{\beta}_{\tau k}(u_i, v_i)$ of each coefficient $\beta_{\tau k}(u_i, v_i)$ ($k = 0, 1, \ldots, p$) therefore can be derived by employing the same approach. It should be noted that bandwidth needs to be determined for local estimations. In a statistical sense, a smaller bandwidth indicates a more local spatial process while a larger bandwidth implies a relative global spatial process (where associations may not vary greatly across space). To find the bandwidth that fits the data best, bandwidth selection is required at each quantile and the task can be conducted with cross-validation method (V. Y.-J. Chen et al. 2012). A series of programs in SAS® or R have been developed to implement bandwidth selection and the GWQR analysis. Other technical details regarding the estimations of standard errors and the significance tests of the local parameters could be found elsewhere (V. Y.-J. Chen et al. 2012).

Assessments for Nonstationarity of Regression Coefficient

According to Fotheringham et al. (2002), assessing whether the regression parameters really vary across space is crucial to validate the use of GWR-type modeling for the investigations of spatially varying relationships. This amounts to conduct statistical hypothesis tests for spatial nonstationarity or inconstancy of regression coefficients. For GWQR, the task in original paper (V. Y.-J. Chen et al. 2012) is accomplished by comparing the interquartile range (IQR) of the local estimates obtained

from GWQR with the standard errors of the global QR estimators. If the IQR is twice as large as the standard error, it indicates the variability of the estimates over space. Such examination is straightforward and similar to the rule of thumb proposed by Fotheringham et al. We acknowledge that a preferable test can be developed with Monte Carlo technique. This technique has been commonly utilized in GWR literature, despite recent studies have begun to explore the use of bootstrap tests for determination of constant coefficients (Li and Mei 2018; Mei et al. 2016). We here conduct the Monte Carlo test with a procedure that randomly permutes geographical coordinates of the observations under the null hypothesis of spatial stationarity based on a total of M replications of GWQR model. More specifically, for each quantile function, we specify the null hypothesis that the local coefficients are equal across locations, i.e., $\beta_{\tau k}(u_i, v_i) = \beta_{\tau k}$. The variance of its estimates $\hat{\beta}_{\tau k}(u_i, v_i)$ is a useful indicator to measure the variability of regression parameters across space. Such variance is firstly obtained based on actual location data, called V_k. Then we permute the locations for $M - 1$ times and compute the variance for each permutation. The total M calculated variances formulate an empirical distribution and hence the ranking of V_k in this distribution, say R, can be observed. The p-value for testing the null hypothesis of spatial stationarity is then approximately by R/M.

IV. Illustration

We illustrate how to use GWQR in population research with the analysis of U.S. county-level COVID-19 vaccination data. The goal is to demonstrate the versatility of GWQR and its applicability to demographic research questions. The county-level data are used in a recent study (Yang et al. 2022) and include data from various federal agencies. Specifically, the

dependent variable of this illustration is the COVID-19 vaccination rate, which is defined as the percentage of population (aged 18 or older) who were fully vaccinated between December 13, 2020, and October 26, 2021. This variable is from the data maintained by the Centers for Disease Control and Prevention (CDC). Regarding independent variables, we include the following six variables in the analysis: percentage of Republican votes in 2020 presidential election, median family income (in natural logarithm), poverty, percentage of non-Hispanic Blacks, percentage of Hispanics, and COVID-19 case rate (total confirmed cases per 1,000 population between January 21, 2020 and October 26, 2021). The presidential election county-level data are publicly available (see https://github.com/tonmcg/US_County_Level_Election_Results_08-20) and the COVID-19 case rates are from CDC. The other four variables are calculated with the 2015-2019 American Community Survey 5-year estimates (see https://www.census.gov/newsroom/press-kits/2020/acs-5-year.html). This illustration focuses on the 3,106 counties in the contiguous U.S. and these independent variables have been identified as important factors for county-level COVID-19 vaccination rate (Yang et al. 2022). The details about variable construction are available upon request and the descriptive statistics of these variables are summarized in Table 1. The last column of Table 1 lists the variance inflation factors (VIF) of the six independent variables. All the VIF values are less than 5, suggesting that multicollinearity is not a concern.

We first fit global models with the technique of QR at the following selected quantiles: 5th, 25th, 50th, 75th, and 95th. The QR results are shown in Table 2 and several findings are notable. First, the association between the COVID-19 case rate and the COVID-19 vaccination rate is positive, and it seems to decrease from low percentiles to high percentiles. Explicitly, at the 5th percentile, a one percentage point change is associated with

Table 1. Descriptive statistics for variables for U.S. COVID-19 vaccination
 ($n = 3,106$)

Variable	Min	Max	Mean	SD	VIF
Adult vaccination rate	0.000	99.900	53.020	14.342	
Case rate	2.710	47.263	14.499	3.715	1.189
Percentage of replication votes	5.397	96.182	65.040	16.069	1.803
Logged median income	9.976	11.866	10.852	0.250	4.232
Poverty rate	2.256	48.222	14.833	5.940	4.253
Percentage of non-Hispanic Blacks	0.000	87.226	9.013	14.488	1.712
Percentage of Hispanics	0.000	99.174	9.459	13.934	1.106

Note: COVID-19: coronavirus disease; VIF: variance inflation factor.

Table 2. QR results for U.S. COVID-19 vaccination

	Percentile				
Variable	5th	25th	50th	75th	95th
Intercept	91.796	59.608***	56.614***	69.088***	80.509
Case rate	1.336***	0.961***	0.647***	0.449***	0.451***
Percentage of replication votes	-0.604***	-0.676***	-0.635***	-0.646***	-0.707***
Logged median income	-2.195	2.655	3.345*	2.860	2.830
Poverty rate	-0.715***	-0.471***	-0.306***	-0.196**	-0.109
Percentage of non-Hispanic Blacks	-0.377***	-0.390***	-0.303***	-0.325***	-0.324***
Percentage of Hispanics	0.129**	0.045***	0.033*	0.017	0.033

Note: QR: quantile regression; COVID-19: coronavirus disease.
*$p < .05$; **$p < .01$; ***$p < .001$.

a 1.336-percentage-point increase in vaccination rate; however, this relationship shrinks by almost two-third to 0.451 at the 95th percentile. This pattern suggests that the relationship between COVID-19 case rate and COVID-19 vaccination rate may be statistically heterogeneous. Poverty also demonstrates a similar pattern. Second, compared with other independent

variables, percentage of non-Hispanic Blacks has a relatively stable relationship with the COVID-19 vaccination rates across the percentiles, ranging between -0.390 and -0.303. Third, median income and percentage of Hispanics seem to have a significant relationship with the lower percentiles of the dependent variable, rather than those higher than the 50th percentile. This, again, indicates statistical heterogeneity.

After obtaining the QR results, the GWQR analysis is conducted. We implement two GWQR models, GWQR-local linear and GWQR-local constant, for the five quantiles. As previously discussed (see Section III), the two models differ in their estimating approaches that approximate the relationships between the independent and the response variables over a set of neighboring observations. The local linear estimator is expected to be more accurate in model predictions due to the high order approximation (N. Wang et al. 2008). We compare the model performance with two goodness-of-fit measures: total sum of check loss and root mean square prediction error (RMSPE). In addition, we investigate whether the model residuals are spatially autocorrelated with Moran's I. As shown in Table 3, compared with QR, both GWQR models fit the vaccination data better and have weaker spatial autocorrelation at each quantile. Also, the local linear GWQR outperforms the local constant GWQR for all percentiles with smaller values of the criteria, consistent with our expectation.

To summarize the massive results generated by GWQR and assist in a comparison of results across space and quantile, at each quantile and for each independent variable, we follow V. Y.-J. Chen et al. (2012, 2020) and provide the five-number summary statistics (minimum, lower quartile, median, upper quartile, and maximum) in Table 4. Some findings can be drawn as follows. First, the bandwidths at 5th and 95th percentiles of the county-level vaccination rates are larger than those of other percentiles. This

Table 3. Comparisons of QR and GWQR for U.S. COVID-19 vaccination

Criterion/model	Percentile				
	5th	25th	50th	75th	95th
Total sum of check loss					
QR	4,349.969	10,054.170	11,323.900	8,693.950	2,936.434
GWQR-local constant	2,262.932	6,663.974	7,985.211	6,275.332	2,438.530
GWQR-local linear	2,209.718	6,374.852	7,576.130	6,265.389	2,258.728
RMSPE					
QR	20.7978	11.5677	10.5904	12.1597	18.0672
GWQR-local constant	16.6104	9.1881	8.0742	9.5606	15.9454
GWQR-local linear	16.0564	8.7945	7.7883	9.4222	15.2150
Moran's I					
QR	0.4278^{***}	0.4208^{***}	0.4221^{***}	0.4219^{***}	0.4142^{***}
GWQR-local constant	0.3835^{***}	0.1498^{***}	0.1626^{***}	0.1925^{***}	0.3174^{***}
GWQR-local linear	0.3676^{***}	0.1189^{***}	0.1431^{***}	0.1905^{***}	0.3080^{***}

Note: QR: quantile regression; GWQR: geographically weighted quantile regression; COVID-19: coronavirus disease; RMSPE: root mean square prediction error.

$^{***} p < .001$.

Table 4. Parameter estimates of local linear GWQR for U.S. COVID-19 vaccination

	Min	Q1	Median	Q3	Max	p-value[a]
Q = 0.05 (bandwidth = 526)						
Intercept	-1,319.080	-132.062	0.742	108.297	876.624	.102
Case rate	-3.423	0.413	0.861	1.346	3.587	.348
Percentage of replication votes	-1.648	-0.733	-0.578	-0.377	1.681	<.001
Logged median income	-69.290	-2.012	6.925	18.096	118.252	.170
Poverty rate	-2.262	-0.839	-0.288	0.233	0.957	.026
Percentage of non-Hispanic Blacks	-7.458	-0.676	-0.328	0.168	3.972	<.001
Percentage of Hispanics	-1.419	-0.101	0.154	0.336	2.673	<.001
Q = 0.25 (bandwidth = 317)						
Intercept	-1,147.127	-60.112	20.916	141.305	751.478	<.001
Case rate	-3.941	0.222	0.642	1.127	2.147	<.001
Percentage of replication votes	-1.275	-0.741	-0.598	-0.472	1.224	<.001
Logged median income	-57.355	-5.109	6.119	13.260	100.731	<.001
Poverty rate	-3.200	-0.737	-0.333	-0.014	5.004	<.001
Percentage of non-Hispanic Blacks	-6.600	-0.556	-0.318	-0.110	1.990	<.001
Percentage of Hispanics	-1.344	-0.179	0.055	0.287	2.202	<.001

Table 4. Parameter estimates of local linear GWQR for U.S. COVID-19 vaccination (continued)

	Min	Q1	Median	Q3	Max	p-value[a]
	Q = 0.50 (bandwidth = 262)					
Intercept	-476.150	-62.039	43.735	154.559	933.312	.002
Case rate	-2.771	0.131	0.479	0.937	2.475	<.001
Percentage of replication votes	-1.040	-0.705	-0.566	-0.448	0.384	<.001
Logged median income	-72.657	-5.806	4.486	13.225	48.726	.002
Poverty rate	-2.420	-0.703	-0.343	-0.001	1.740	.002
Percentage of non-Hispanic Blacks	-6.226	-0.588	-0.309	-0.083	1.742	<.001
Percentage of Hispanics	-1.387	-0.150	0.044	0.256	3.095	<.001
	Q = 0.75 (bandwidth = 368)					
Intercept	-354.871	-40.814	53.421	166.673	725.455	<.001
Case rate	-2.816	-0.002	0.293	0.706	2.228	<.001
Percentage of replication votes	-1.070	-0.669	-0.568	-0.420	0.580	<.001
Logged median income	-51.775	-5.698	3.886	11.941	38.078	<.001
Poverty rate	-2.388	-0.786	-0.414	-0.044	2.074	<.001
Percentage of non-Hispanic Blacks	-3.910	-0.609	-0.314	-0.096	0.703	<.001
Percentage of Hispanics	-0.868	-0.102	0.018	0.213	1.882	<.001

Table 4. Parameter estimates of local linear GWQR for U.S. COVID-19 vaccination (continued)

	Min	Q1	Median	Q3	Max	p-value[a]
	Q = 0.95 (bandwidth = 857)					
Intercept	-384.055	-2.245	89.532	154.651	787.831	.050
Case rate	-2.189	-0.172	0.116	0.495	1.923	.008
Percentage of replication votes	-1.333	-0.722	-0.588	-0.442	0.391	< .001
Logged median income	-56.596	-3.969	1.680	8.987	43.805	.044
Poverty rate	-2.861	-0.765	-0.422	-0.105	1.183	.008
Percentage of non-Hispanic Blacks	-6.396	-0.642	-0.324	-0.021	1.108	< .001
Percentage of Hispanics	-0.809	-0.291	-0.062	0.084	0.951	< .001

Note: GWQR: geographically weighted quantile regression; COVID-19: coronavirus disease.

[a] The p-value is of Monte Carlo test.

bandwidth pattern is consistent with what has been found in V. Y.-J. Chen et al. (2020). The largest bandwidth in Table 4 was 857, which is roughly 28 percent of the total observations of counties ($n = 3,106$). From the statistical modeling perspective, the low proportion indicates that the spatial processes between vaccination and the independent variables tend to operate at the local, rather than global, level. Second, the Monte Carlo tests (the last column of Table 4) based on 499 permutations report the presence of spatial heterogeneity in the regression relationships at most quantiles, except for the lowest one (5th) where the associations of some covariates are constant. The results imply that spatial heterogeneity should not be overlooked in this U.S. county COVID-19 vaccination dataset. Third, the associations between the explanatory variables and vaccination rates also vary across quantiles. For instance, the magnitude of the estimated relationship of vaccination with poverty tends to be stronger in the higher percentiles than in the lower ones, and the magnitude of the relationship of COVID-19 case rate with vaccination rate steadily decreases from the low to the high end of the vaccination distribution.

In order to better interpret both heterogeneities in the GWQR results, we further visualize the local estimates with the mapping technique proposed in Matthews and Yang (2012). Due to space limitations and also for the illustrative intent, we focus our discussion on a key independent variable that is not explicitly discussed by Yang et al. (2022). Detailed results of the GWQR analysis are available upon request.

Figure 1 demonstrates the heterogeneous associations between the COVID-19 case rate and the COVID-19 vaccination rate across quantiles. The colored areas are the areas where the associations are statistically significant at the 5% of significance level. Three spatial patterns are worth noting. First, counties in southern states, especially Florida, Georgia, and
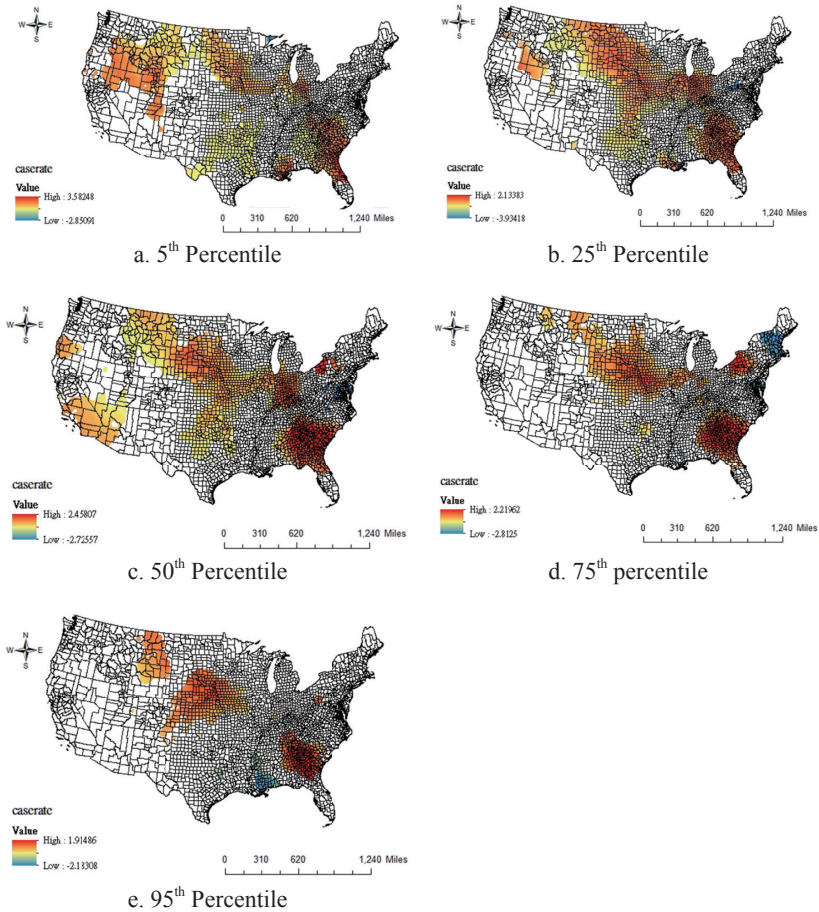
a. 5th Percentile

b. 25th Percentile

c. 50th Percentile

d. 75th percentile

e. 95th Percentile

Figure 1. Heterogeneous associations between the case rate and vaccination by quantiles

Note: See the online version (https://doi.org/10.6191/JPS) for the full-colored figure.

Alabama, consistently show a positive relationship between case rate and vaccination rate. This pattern indicates that regardless of the absolute level of vaccination rate, higher COVID-19 case rates may pose a threat to population health, which in turn encourage population aged 18 and over to be fully vaccinated. Second, counties of several states in upper Great

Plains, such as North Dakota, South Dakota, and Nebraska, show a positive relationship between case rate and vaccination rate, and these patterns are largely consistently across percentiles. Third, few counties are estimated to have a negative relationship between case rate and vaccination rate, and these counties are observed in the northeastern region in higher percentiles (e.g., 75th percentile). These maps highlight spatial heterogeneity in not only the significance of local estimates but also the direction of estimated relationship between COVID-19 case rate and vaccination rate in U.S. counties.

V. Discussions and Conclusions

The past few decades have witnessed rapid changes in spatial demography with respect to methodological development, data collection, and complex dynamics between people and places (Matthews et al. 2021). As demographic research heavily relies on the analysis of data that are across a range of spatial scale, it becomes a challenge to develop new method that may produce insights into existing literature (Matthews et al. 2012). The importance of spatial structure, spatial dependence, and spatial heterogeneity is increasingly emphasized and considered in population research (Assunção et al. 2005; Schmertmann et al. 2008; Sparks et al. 2013). However, demographers mainly concentrate on the commonly used spatial methods developed to address spatial homogeneity (Chi and Zhu 2008; Voss 2007), such as spatial autoregressive regression and global spatial econometrics models. By contrast, relatively little attention has been paid to heterogeneity, both spatial and statistical.

The purpose of the case study is to verify the applicability of the GWQR method, specifically as it pertains to subjects of interest to demographers. In our example that used the data published by Yang et al. (2022), both

spatial and statistical heterogeneity are explored with GWQR. The results suggest that statistical heterogeneity may provide new information relevant to questions across the range of a dependent variable's distribution (i.e., vaccination). Exploring such statistically heterogeneous associations can help demographers understand the geographical differentials and better investigate the factors that may improve vaccination in counties where the vaccination rate is low. Spatial heterogeneity also appears to commonly exist in our illustration. As shown in Figure 1, the COVID-19 case rates have shown a spatially heterogeneous relationship with the COVID-19 vaccination rates in U.S. counties across the five percentiles in that many counties show a null relationship but others demonstrate either a positive or a negative association. As such, it should be emphasized that these heterogeneities may not be as rare as we might think in demographic studies. The GWQR analysis challenges a global model perspective (i.e., QR model) that is unable to identify spatial heterogeneity (V. Y.-J. Chen et al. 2020) and perhaps uncover new knowledge. For example, without the GWQR approach, the finding of a negative relationship between case rate and vaccination rate in some northeastern counties would not have been identified.

Though the value of GWQR has been validated, several issues warrant further attention. First, the larger bandwidths at the extreme quantiles could be attributed to the loss function of GWQR as pointed by V. Y.-J. Chen et al. (2020). This suggests that the GWQR needs to draw on more observations in the kernel density function for reliable parameter estimations at the extremes. By contrast, based on the empirical viewpoint, a large bandwidth indicates that the local model results are close to those obtained from the global model. Second, this study draws inferential properties of GWQR using the asymptotic theory developed in V. Y.-J. Chen et al. (2012) for the empirical application. While the asymptotic inference

may suffer computation difficulties (V. Y.-J. Chen et al. 2020), we note that a bootstrapping approach proposed in V. Y.-J. Chen et al. (2020) can be used as an alternative inferential framework for GWQR. Other technical concerns regarding variable selection of GWQR or the possibility of spatial autoregression in the response variable with GWQR have been also studied in W. Wang et al. (2018) and Tomal and Helbich (2023). On the other hand, a sophisticated statistical test for the hypothesis that local estimates are equal across quantiles should be developed in the future. Bootstrap method may also be useful to address such limitation. Furthermore, the analysis results of the 5th percentile in our example indicate that some regression coefficients are spatially invariant. A semiparametric or mixed GWQR allowing for global impacts in the GWQR model specification is desirable. Finally, using the five-number summary statistics and visualizing spatial heterogeneity are appropriate techniques to synthesize the massive amount of information generated by GWQR. How to fully utilize all of the results to simultaneously present both statistical and spatial heterogeneity remains a challenge.

To promote GWQR in population research, we would like to offer some suggestions as to when and why GWQR might be used to draw insight from heterogeneities. First, when the spatial information (e.g., coordinates) of the observations is available and the literature has assumed spatial and statistical homogeneity between the dependent and independent variables, researchers may want to adopt GWQR to test both homogeneity assumptions. Exploring the existence of heterogeneity may shed new light into population- and/or place-specific policies that could effectively address social and demographic problems (Han et al. 2011; Partridge and Rickman 2007; Shoff et al. 2012). Second, when the mean or median value of the dependent variable is not of primary interest and the dependent variable

is unevenly distributed across space, GWQR might help to untangle the statistically and spatially heterogeneous relationships. For instance, the second demographic transition (SDT) theory has discovered distinctive spatial patterns in the U.S. (Lesthaeghe and Neidert 2006). Should the impact of SDT on low infant birth weight be studied, a GWQR approach may be useful. Finally, when researchers would like to interpolate the associations between the dependent and independent variables across space and/or the distribution of the dependent variable, GWQR, along with other spatial QR methods (Lum and Gelfand 2012; Reich et al., 2011), is an appropriate tool.

To sum up, spatial demography often focuses on exploring demographic processes in space using spatial analytic methods (Weeks 2004). Though spatial modeling techniques have been considered in population studies, spatial and statistical heterogeneity have not yet received much attention and discussed simultaneously. This paper directly addresses such concern by demonstrating the use of GWQR for exploring heterogeneities in population research. Demography has a long history of investigating population dynamics and space (Matthews and Parker 2013; Matthews et al. 2021). We believe a focus on heterogeneities will help move the field forward.

References

Abrevaya, J. and C. M. Dahl. 2008. "The Effects of Birth Inputs on Birthweight: Evidence from Quantile Estimation on Panel Data." *Journal of Business & Economic Statistics* 26(4): 379-397. doi:10.1198/073500107000000269

Anselin, L. 1988a. "Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity." *Geographical Analysis* 20(1): 1-17. doi:10.1111/j.1538-4632.1988.tb00159.x

Anselin, L. 1988b. *Spatial Econometrics: Methods and Models*. Dordrecht, The Netherlands: Kluwer Academic.

Assunção, R. M., C. P. Schmertmann, J. E. Potter, and S. M. Cavenaghi. 2005. "Empirical Bayes Estimation of Demographic Schedules for Small Areas." *Demography* 42(3): 537-558. doi:10.1353/dem.2005.0022

Austin, P. C., J. V. Tu, P. A. Daly, and D. A. Alter. 2005. "The Use of Quantile Regression in Health Care Research: A Case Study Examining Gender Differences in the Timeliness of Thrombolytic Therapy." *Statistics in Medicine* 24(5): 791-816. doi:10.1002/sim.1851

Brunsdon, C., A. S. Fotheringham, and M. Charlton. 1998. "Geographically Weighted Regression—Modelling Spatial NonStationarity." *Journal of the Royal Statistical Society: Series D (The Statistician)* 47(3): 431-443. doi:10.1111/1467-9884.00145

Chen, C. and Y. Wei. 2005. "Computational Issues for Quantile Regression." *Sankhyā: The Indian Journal of Statistics* 67(2): 399-417.

Chen, C.-M., C.-K. Chang, and C.-Y. Yeh. 2012. "A Quantile Regression Approach to Re-Investigate the Relationship between Sleep Duration

and Body Mass Index in Taiwan." *International Journal of Public Health* 57: 485-493. doi:10.1007/s00038-011-0239-7

Chen, V. Y.-J., W.-S. Deng, T.-C. Yang, and S. A. Matthews. 2012. "Geographically Weighted Quantile Regression (GWQR): An Application to U.S. Mortality Data." *Geographical Analysis* 44(2): 134-150. doi:10.1111/j.1538-4632.2012.00841.x

Chen, V. Y.-J., K. Park, F. Sun, and T.-C. Yang. 2022. "Assessing COVID-19 Risk with Temporal Indices and Geographically Weighted Ordinal Logistic Regression in US Counties." *PLOS ONE* 17(4): e0265673. doi:10.1371/journal.pone.0265673

Chen, V. Y.-J., P.-C. Wu, T.-C. Yang, and H.-J. Su. 2010. "Examining Non-Stationary Effects of Social Determinants on Cardiovascular Mortality after Cold Surges in Taiwan." *Science of the Total Environment* 408(9): 2042-2049. doi:10.1016/j.scitotenv.2009.11.044

Chen, V. Y.-J. and T.-C. Yang. 2012. "SAS Macro Programs for Geographically Weighted Generalized Linear Modeling with Spatial Point Data: Applications to Health Research." *Computer Methods and Programs in Biomedicine* 107(2): 262-273. doi:10.1016/j.cmpb.2011.10.006

Chen, V. Y.-J., T.-C. Yang, and S. A. Matthews. 2020. "Exploring Heterogeneities with Geographically Weighted Quantile Regression: An Enhancement Based on The Bootstrap Approach." *Geographical Analysis* 52(4): 642-661. doi:10.1111/gean.12229

Chi, G. and J. Zhu. 2008. "Spatial Regression Models for Demographic Analysis." *Population Research and Policy Review* 27(1): 17-42. doi:10.1007/s11113-007-9051-8

Cressie, N. A. C. 1993. *Statistics for Spatial Data*, Rev. ed. New York, NY: Wiley.

Drewnowski, A., C. D. Rehm, and D. Solet. 2007. "Disparities in Obesity Rates: Analysis by ZIP Code Area." *Social Science & Medicine* 65(12): 2458-2463. doi:10.1016/j.socscimed.2007.07.001

Dutilleul, P. and P. Legendre. 1993. "Spatial Heterogeneity against Heteroscedasticity: An Ecological Paradigm versus a Statistical Concept." *Oikos* 66(1): 152-171. doi:10.2307/3545210

Finley, A. O. 2011. "Comparing Spatially-Varying Coefficients Models for Analysis of Ecological Data with Non-Stationary and Anisotropic Residual Dependence." *Methods in Ecology and Evolution* 2(2): 143-154. doi:10.1111/j.2041-210X.2010.00060.x

Fotheringham, A. S. 1997. "Trends in Quantitative Methods I: Stressing the Local." *Progress in Human Geography* 21(1): 88-96. doi:10.1191/030913297676693207

Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationship*. New York, NY: Wiley.

Gebreab, S. Y. and A. V. Diez Roux. 2012. "Exploring Racial Disparities in CHD Mortality between Blacks and Whites across the United States: A Geographically Weighted Regression Approach." *Health & Place* 18(5): 1006-1014. doi:10.1016/j.healthplace.2012.06.006

Gelfand, A. E., H.-J. Kim, C. F. Sirmans, and S. Banerjee. 2003. "Spatial Modeling with Spatially Varying Coefficient Processes." *Journal of the American Statistical Association* 98(462): 387-396. doi:10.1198/016214503000170

Glaser, R. E. 1983. "Homogeneity and Tests of Homogeneity." Pp. 662-665 in *Encyclopedia of Statistical Sciences* (Vol. 3), edited by S. Kotz and N. L. Johnson. New York, NY: Wiley.

Gutiérrez-Posada, D., F. Rubiera-Morollon, and A. Viñuela. 2017.

"Heterogeneity in the Determinants of Population Growth at the Local Level: Analysis of the Spanish Case with a GWR Approach." *International Regional Science Review* 40(3): 211-240. doi:10.1177/0160017615589009

Han, E., L. M. Powell, and O. Pugach. 2011. "The Heterogenous Relationship of Food Stamp Participation with Body Mass: Quantile Regression Model." *Food Policy* 36(4): 499-506. doi:10.1016/j.foodpol.2011.04.004

Hao, L. and D. Q. Naiman. 2007. *Quantile Regression*. Thousand Oaks, CA: Sage.

Harris, P. 2019. "A Simulation Study on Specifying a Regression Model for Spatial Data: Choosing between Autocorrelation and Heterogeneity Effects." *Geographical Analysis* 51(2): 151-181. doi:10.1111/gean.12163

Kleinbaum, D. G., L. L. Kupper, A. Nizam, and K. E. Muller. 2008. *Applied Regression Analysis and Other Multivariable Methods*, 4th ed. Belmont, CA: Duxbury Press.

Koenker, R. 2000. "Galton, Edgeworth, Frisch, and Prospects for Quantile Regression in Econometrics." *Journal of Econometrics* 95(2): 347-374. doi:10.1016/S0304-4076(99)00043-3

Koenker, R. 2005. *Quantile Regression*. Cambridge, UK: Cambridge University Press.

Koenker, R. and G. Bassett, Jr. 1978. "Regression Quantiles." *Econometrica* 46(1): 33-50. doi:10.2307/1913643

LeSage, J. and R. K. Pace. 2009. *Introduction to Spatial Econometrics*. Boca Raton, FL: Chapman and Hall/CRC.

Lesthaeghe, R. J. and L. Neidert. 2006. "The Second Demographic Transition in the United States: Exception or Textbook Example?"

Population and Development Review 32(4): 669-698. doi:10.1111/ j.1728-4457.2006.00146.x

Li, D. and C. Mei. 2018. "A Two-Stage Estimation Method with Bootstrap Inference for Semi-Parametric Geographically Weighted Generalized Linear Models." *International Journal of Geographical Information Science* 32(9): 1860-1883. doi:10.1080/13658816.2018.1463443

Lloyd, C. D. 2010. *Local Models for Spatial Analysis*, 2nd ed. Boca Raton, FL: Chapman and Hall/CRC.

Lum, K. and A. E. Gelfand. 2012. "Spatial Quantile Multiple Regression Using the Asymmetric Laplace Process." *Bayesian Analysis* 7(2): 235-258. doi:10.1214/12-BA708

Machado, J. A. F. and J. Mata. 2005. "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression." *Journal of Applied Econometrics* 20(4): 445-465. doi:10.1002/jae.788

Matthews, S. A. 2011. "Spatial Polygamy and the Heterogeneity of Place: Studying People and Place via Egocentric Methods." Pp. 35-55 in *Communities, Neighborhoods, and Health: Expanding the Boundaries of Place*, edited by L. M. Burton, S. P. Kemp, M. Leung, S. A. Matthews, and D. T. Takeuchi. New York, NY: Springer. doi:10.1007/978-1-4419-7482-2_3

Matthews, S. A., D. G. Janelle, and M. F. Goodchild. 2012. "Future Directions in Spatial Demography Specialist Meeting: Final Report." https://escholarship.org/content/qt7xx3k2z4/qt7xx3k2z4.pdf?t=nk8ydx (Date visited: May 17, 2023).

Matthews, S. A. and D. M. Parker. 2013. "Progress in Spatial Demography." *Demographic Research* 28: 271-312. doi:10.4054/DemRes.2013.28.10

Matthews, S. A., L. Stiberman, J. Raymer, T.-C. Yang, E. Gayawan, S. Saita, ... and D. W. S. Wong. 2021. "Looking Back, Looking Forward:

Progress and Prospect for Spatial Demography." *Spatial Demography* 9(1): 1-29. doi:10.1007/s40980-021-00084-9

Matthews, S. A. and T.-C. Yang. 2012. "Mapping the Results of Local Statistics: Using Geographically Weighted Regression." *Demographic Research* 26: 151-166. doi:10.4054/DemRes.2012.26.6

Mei, C.-L., M. Xu, and N. Wang. 2016. "A Bootstrap Test for Constant Coefficients in Geographically Weighted Regression Models." *International Journal of Geographical Information Science* 30(8): 1622-1643. doi:10.1080/13658816.2016.1149181

Mollalo, A., B. Vahedi, and K. M. Rivera. 2020. "GIS-Based Spatial Modeling of COVID-19 Incidence Rate in the Continental United States." *Science of The Total Environment* 728: 138884. doi:10.1016/j.scitotenv.2020.138884

Nakaya, T., A. S. Fotheringham, C. Brunsdon, and M. Charlton. 2005. "Geographically Weighted Poisson Regression for Disease Association Mapping." *Statistics in Medicine* 24(17): 2695-2717. doi:10.1002/sim.2129

Páez, A., S. Farber, and D. Wheeler. 2011. "A Simulation-Based Study of Geographically Weighted Regression as a Method for Investigating Spatially Varying Relationships." *Environment and Planning A: Economy and Space* 43(12): 2992-3010. doi:10.1068/a44111

Partridge, M. D. and D. S. Rickman. 2007. "Persistent Pockets of Extreme American Poverty and Job Growth: Is There a Place-Based Policy Role?" *Journal of Agricultural and Resource Economics* 32(1): 201-224. doi:10.22004/ag.econ.8599

Partridge, M. D., D. S. Rickman, M. R. Olfert, and K. Ali. 2012. "Dwindling U.S. Internal Migration: Evidence of Spatial Equilibrium or Structural Shifts in Local Labor Markets?" *Regional Science & Urban Economics*

42(1-2): 375-388. doi:10.1016/j.regsciurbeco.2011.10.006

Pickett, S. T. A. and M. L. Cadenasso. 1995. "Landscape Ecology: Spatial Heterogeneity in Ecological Systems." *Science* 269(5222): 331-334. doi:10.1126/science.269.5222.331

Raudenbush, S. W. and A. S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd ed. Thousand Oaks, CA: Sage.

Reich, B. J., M. Fuentes, and D. B. Dunson. 2011. "Bayesian Spatial Quantile Regression." *Journal of the American Statistical Association* 106(493): 6-20. doi:10.1198/jasa.2010.ap09237

Schmertmann, C. P., R. M. Assunção, and J. E. Potter. 2010. "Knox Meets Cox: Adapting Epidemiological Space–Time Statistics to Demographic Studies." *Demography* 47(3): 629-650. doi:10.1353/dem.0.0113

Schmertmann, C. P., J. E. Potter, and S. M. Cavenaghi. 2008. "Exploratory Analysis of Spatial Patterns in Brazil's Fertility Transition." *Population Research and Policy Review* 27(1): 1-15. doi:10.1007/s11113-007-9052-7

Shoff, C., V. Y.-J. Chen, and T.-C. Yang. 2014. "When Homogeneity Meets Heterogeneity: The Geographically Weighted Regression with Spatial Lag Approach to Prenatal Care Utilisation." *Geospatial Health* 8(2): 557-568. doi:10.4081/gh.2014.45

Shoff, C., T. C. Yang, and S. A. Matthews. 2012. "What Has Geography Got to Do with It? Using GWR to Explore Place-Specific Associations with Prenatal Care Utilization." *GeoJournal* 77(3): 331-341. doi:10.1007/s10708-010-9405-3

Sokal, R. R. and F. J. Rohlf. 1981. *Biometry: The Princeples and Practice of Statistics in Biological Research*, 2nd ed. San Francisco, CA: W. H. Freeman and Company.

Sparks, J., C. S. Sparks, and J. J. A. Campbell. 2013. "An Application of Bayesian Spatial Statistical Methods to the Study of Racial and Poverty Segregation and Infant Mortality Rates in the US." *GeoJournal* 78(2): 389-405. doi:10.1007/s10708-011-9445-3

Tobler, W. R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46(Supplement 1): 234-240. doi:10.2307/143141

Tomal, M. and M. Helbich. 2023. "A Spatial Autoregressive Geographically Weighted Quantile Regression to Explore Housing Rent Determinants in Amsterdam and Warsaw." *Environment and Planning B: Urban Analytics and City Science* 50(3): 579-599 doi:10.1177/23998083221122790

Voss, P. R. 2007. "Demography as a Spatial Social Science." *Population Research and Policy Review* 26(5): 457-476. doi:10.1007/s11113-007-9047-4

Voss, P. R., D. D. Long, R. B. Hammer, and S. Friedman. 2006. "County Child Poverty Rates in the US: A Spatial Regression Approach." *Population Research and Policy Review* 25(4): 369-391. doi:10.1007/s11113-006-9007-4

Waller, L. A., L. Zhu, C. A. Gotway, D. M. Gorman, and P. J. Gruenewald. 2007. "Quantifying Geographic Variations in Associations between Alcohol Distribution and Violence: A Comparison of Geographically Weighted Regression and Spatially Varying Coefficient Models." *Stochastic Environmental Research and Risk Assessment* 21(5): 573-588. doi:10.1007/s00477-007-0139-9

Wang, N., C.-L. Mei, and X.-D. Yan. 2008. "Local Linear Estimation of Spatially Varying Coefficient Models: An Improvement on the Geographically Weighted Regression Technique." *Environment and Planning A: Economy and Space* 40(4): 986-1005. doi:10.1068/a3941

Wang, W., S. Xu, and T. Yan. 2018. "Structure Identification and Model Selection in Geographically Weighted Quantile Regression Models." *Spatial Statistics* 26: 21-37. doi:10.1016/j.spasta.2018.05.003

Weeks, J. R. 2004. "The Role of Spatial Analysis in Demographic Research." Pp. 381-399 in *Spatially Integrated Social Science*, edited by M. F. Goodchild and D. G. Janelle. New York, NY: Oxford University Press.

Wheeler, D. C. and C. A. Calder. 2007. "An Assessment of Coefficient Accuracy in Linear Regression Models with Spatially Varying Coefficients." *Journal of Geographical Systems* 9(2): 145-166. doi:10.1007/s10109-006-0040-y

Wheeler, D. C. and A. Páez. 2010. "Geographically Weighted Regression." Pp. 461-486 in *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, edited by M. M. Fischer and A. Getis. Heidelberg, Germany: Springer. doi:10.1007/978-3-642-03647-7_22

Wheeler, D. C. and M. Tiefelsdorf. 2005. "Multicollinearity and Correlation among Local Regression Coefficients in Geographically Weighted Regression." *Journal of Geographical Systems* 7(2): 161-187. doi:10.1007/s10109-005-0155-6

Wheeler, D. C. and L. A. Waller. 2009. "Comparing Spatially Varying Coefficient Models: A Case Study Examining Violent Crime Rates and Their Relationships to Alcohol Outlets and Illegal Drug Arrests." *Journal of Geographical Systems* 11(1): 1-22. doi:10.1007/s10109-008-0073-5

Yang, T.-C., V. Y.-J. Chen, C. Shoff, and S. A. Matthews. 2012. "Using Quantile Regression to Examine the Effects of Inequality Across the Mortality Distribution in the U.S. Counties." *Social Science & Medicine* 74(12): 1900-1910. doi:10.1016/j.socscimed.2012.02.029

Yang, T.-C., S. A. Matthews, and F. Sun. 2022. "Multiscale Dimensions of Spatial Process: COVID-19 Fully Vaccinated Rates in U.S. Counties." *American Journal of Preventive Medicine* 63(6): 954-961. doi:10.1016/ j.amepre.2022.06.006

以地理加權分量迴歸探討人口研究之空間和統計異質性

陳怡如[*]　楊澤全[**]

摘要

　　愈來愈多人對於探索自變數與反應變數間在不同分量下（使用分量迴歸〔quantile regression〕）或在空間上（使用地理加權迴歸〔geographically weighted regression〕）之異質關係感到興趣。前者通常被視為一種統計異質性，而後者則指的是空間異質性。然而，人口研究在採用這些方法方面進展略顯緩慢。本研究首先簡要討論了為何人口研究者需要多加關注異質性的概念，並介紹了一種同時考量統計和空間異質性的方法，即地理加權分量迴歸（geographically weighted quantile regression, GWQR）。我們以美國各郡新型冠狀病毒（coronavirus disease, COVID-19）疫苗接種率之資料為例，說明如何應用 GWQR，並解釋 GWQR 如何分析疫苗接種率及其因子之間在空間和疫苗接種率分布上的異質性。研究結果顯示，空間和統計異質性是普遍存在的。例如，COVID-19 病例率在位於疫苗接種率低分位數的郡中其關聯性比在高分位數的郡中更強。這種關聯性的空間分布集中

* 國立政治大學統計學系副教授，通訊作者
　E-mail: viyjchen@nccu.edu.tw
** 紐約州立大學奧爾巴尼分校社會學系副教授
　E-mail: tyang3@albany.edu

在山區州的郡,並轉移到中西部地區。因此,我們認為人口研究應考
慮這兩種異質性。

關鍵詞:異質性、同質性、地理加權迴歸、分量迴歸、空間人口學