# A Centered Index of Spatial Concentration: Expected Influence Approach

Filipe Campante, Quoc-Anh Do

# A Centered Index of Spatial Concentration:
## Expected Influence Approach
## and Application to Population and Capital Cities[*]

Filipe R. Campante[†]and Quoc-Anh Do[‡]

This version: April 2010

### Abstract

We construct a general axiomatic approach to measuring spatial concentration around a center or capital point of interest, a concept with wide applicability from urban economics, economic geography and trade, to political economy and industrial organization. By analogy with expected utility theory, we propose a basic axiom of independence (subgroup consistency) and continuity for a concentration order that ranks any two distributions relative to the capital point. We show that this axiom implies an *expected influence* representation of that order, conceptualizing concentration as an aggregation of the expected influence exerted by the capital on all points in the relevant space (or vice-versa). We then propose two axioms (monotonicity and rank invariance) and prove that they imply that the associated influence function must be a decreasing isoelastic function of the distance to the capital. We apply our index to measure the concentration of population around capital cities across countries and US states, and also in US metropolitan areas. We show its advantages over alternative measures, and explore its correlations with many economic and political variables of interest.

---

# 1 Introduction

Spatial concentration is a very important concept in the social sciences, and in economics in particular – both in the sense of geographical space, as studied by urban economics, economic geography or international trade, and in more abstract settings (e.g. product or policy spaces) that are studied in many different fields, from industrial organization to political economy. As a result, a number of methods have been developed to measure this concept, from relatively *ad hoc* measures such as the Herfindahl index to theoretically grounded approaches such as the "dartboard" method of Ellison and Glaeser (1997), and also including the adaptation of indices used to capture related concepts such as inequality (Gini coefficient, entropy measures). These measures are well-suited to analyzing the concentration of a given variable over a "uniform" space, in which no point is considered to be of particular importance in an *ex ante* sense.

In practice, however, it is often the case that some points are indeed more important than others. In other words, we might be interested in measuring the concentration of a given variable *around* a point (e.g. a city or a specific site), rather than its concentration *over* some area (e.g. a region or country). The standard indices of concentration are not suited to capture this type of situation, as they leave aside plenty of information on actual spatial distributions.

This paper presents a coherent framework to understand the concept of concentration around a capital point of interest across a broad range of applications – ultimately, the concentration of any variable in very general spaces of economic interest. We conceptualize a spatial distribution as describing the probability of an individual observation being located at any given point in the relevant space. This analogy with probability distributions leads us to reach for the tools of expected utility to build an axiomatic *Expected Influence* framework for concentration around a capital point. We then develop this framework to generate a theoretically grounded measure to quantify the concept: a *centered* index of spatial concentration (CISC).

Our approach establishes ordinal properties that should be satisfied by any relation $\succeq_{\mathbf{C}}$ designed to capture the concept of concentration around a capital point $\mathbf{C}$. The first basic axiom builds on the analogy with expected utility theory to pose properties of *Independence*, or *Subgroup Consistency*, and *Continuity* (Axiom 1) that are appealing in light of the probabilistic interpretation underlying the approach – and find a natural counterpart in the literature on the measurement of inequality and poverty. These axioms yield the *Expected Influence Theorem*

(Theorem 1), whereby any concentration order $\succeq_{\mathbf{C}}$ has an Expected Influence (EI) representation. In words, we show that we can understand concentration as an aggregation of the expected influence exerted by the capital on all points in the relevant space (or vice-versa).

We then proceed to give further content to the concentration order $\succeq_{\mathbf{C}}$ by specifying two basic properties any such order should satisfy. *Monotonicity* (Axiom 2) considers a pair of distributions such that one of them, for any given distance to $\mathbf{C}$, places more mass closer to that point than the other, and requires that the former should be ranked as more concentrated around $\mathbf{C}$ than the latter. *Rank Invariance* (Axiom 3) prescribes that the ranking between different distributions is preserved when the unit, or scale, of distance measure is arbitrarily changed – in other words, the ranking of two distributions should not change based on whether distances are measured in miles, kilometers, or millimeters. These two very natural properties, when appended to the EI representation, define the class of CISC (Centered Index of Spatial Concentration) (Theorem 2): the expected influence, with the influence function being a *monotonically decreasing, isoelastic ("constant relative risk aversion") function of the distance to the capital point* $\mathbf{C}$.

We then provide further discussion on how to calibrate the crucial degree of freedom left by the CISC – the elasticity parameter. We can interpret it as measuring how marginal influence is affected by the distance to the capital point, and as measuring how the concentration order reacts to mean-preserving spreads (generalized to many dimensions). These two interpretations define two special cases: the *Linear* CISC (L-CISC), with constant marginal influence, and the *Gravity-based* CISC (G-CISC), which is invariant with respect to uniform mean-preserving spreads and can thus be interpreted as eliciting the "gravitational pull" exerted by the capital. We also discuss how the CISC can be normalized to suit different applications.

Examples of circumstances in which there is specific *ex ante* knowledge of the importance of a given capital point are not hard to come by. Topics in urban economics (such as the study of urban sprawl and urbanization, e.g. Henderson 2003, Glaeser and Kahn 2004), political economy (political importance of capital cities and urban centers, e.g. Ades and Glaeser 1995, Traugott 1995, Campante and Do 2007), international trade (gravity equations, as formalized by Anderson and van Wincoop 2003),[1] industrial organization (concentration of competitors

---

[1]Their formalization involves the concept of multilateral resistance, expressible as a measure of how remote a particular country is from the ensemble of other countries. The geographical concentration of the world around each country, in this sense, is theoretically expected and empirically verified to affect trade flows in and out

around a firm or producer), economic geography ("market potential", e.g. Fujita et al. 1999) – all of these place emphasis on the concentration of population and economic activity around a geographical center of interest. The concept is also important in non-geographical contexts. For instance, one can think about concentration around points of interest within a product space, with obvious applications in IO models of spatial competition, but also in development.[2] The same goes for abstract policy spaces, in political economy.[3]

Not surprisingly in light of that, the literature has grappled with the question of devising a centered measure of spatial concentration. Our EI approach, besides generating a specific family of indices, enables us to systematically evaluate those alternative measures within a unified framework, in very general spaces. We show that measures such as "capital primacy" (e.g. the share of population of a country that lives in the capital or main city, as in Ades and Glaeser 1995 or Henderson 2003) violate continuity and monotonicity, as they discard a lot of information by attaching zero weight to all observations falling outside of the designated boundary. Other approaches, such as the negative exponential density functions from the urban economics literature that has tried to measure the "centrality" of "mononuclear" urban areas, can be shown to violate rank invariance. Another interesting example also comes from the same literature on urban sprawl. Galster et al. (2001), for instance, measure this centrality by the inverse of the sum of the distances of each observation to the central business district. We can show that this is a monotonic transformation of L-CISC, and therefore satisfies all of our axioms. This underscores that our general framework can go much further than the typically *ad hoc* approaches in the literature, which are inherently limited by what an intuitive grasp of the properties of a given space will provide.

The EI approach also relates very naturally to the literature on the measurement of riskiness (Aumann and Serrano 2008) (as is clear from the connection with expected utility theory), inequality, and polarization (Duclos et al. 2004). It provides a foundation to extend these other

---

of that country. In an earlier version of this paper, available upon request, we detail the surprisingly close connection between the measure of concentration we develop and the formula of multilateral resistance.

[2]Hidalgo et al. (2007), for instance, are interested in a product space in which distances measure the likelihood that a country might move from one type of product to another. One might then be interested in how concentrated a country's economy is around a specific industry, say, oil production.

[3]See for instance Baron and Diermeier (2001), where the status quo policy has special clout, and hence the concentration of preferences around that status quo point may be of particular interest. Empirically, one could immediately connect this to the voting records of politicians, or to the collection of opinions from, say, the World Values Surveys.

contexts to multidimensional settings, and also lends itself to the development of a measure of (non-centered) concentration.

The second part of the paper provides an example of empirical implementation of our measure, by computing an index of population concentration around capital cities across countries. We use the L-CISC and G-CISC as illustration, and show that our index provides a much more sensible ranking of countries than currently used *ad hoc* alternatives and non-centered proxies. It also uncovers a negative correlation between the size of population and its concentration around the capital city that is not detected by those alternatives.

In addition, motivated by the idea that political influence diminishes with distance to the capital – as put by Ades and Glaeser (1995, p. 198-199), "spatial proximity to power increases political influence" – we consider the correlation between population concentration and a number of measures of quality of governance. We show that there is a positive correlation between concentration and the checks that are faced by governments, and that this correlation is present only in non-democratic countries.[4] The statistical significance of this correlation is substantially improved by using our index instead of the *ad hoc* alternatives.

We also illustrate how our index can shed light on the issue of the choice of where to locate the capital city, which goes back at least to James Madison during the debates at the US Constitutional Convention of 1787. We show that there is a pattern in which both very autocratic and very democratic countries tend to have their capital cities in places with relatively low concentration of population. Inspired by the Madisonian origins of this debate, we extend our implementation by computing our index for US states, and finish it off by running the computations for US metropolitan areas.

The remainder of the paper is organized as follows. Section 2 presents the main definitions required by the theory. Section 3 presents the foundations of the EI approach, and Section 4 adds the axioms that characterize the CISC. Section 5 discusses the interpretation of the elasticity parameter, normalization procedures, and the comparison with other measures. Section 6 contains the empirical implementation and correlation analysis, and Section 7 concludes.

---

[4]This is consistent with Campante and Do (2007), who present a theory of revolutions and redistribution where concentration is key in increasing the redistributive pressures faced by non-democratic governments.

# 2 Main Definitions

We start by spelling out the definitions of the main mathematical objects and transformations that are required for our approach. Our main concern is the spatial concentration of a variable – which can be thought of as population, economic activity, etc. – around a point of interest, so we consider a point $\mathbf{C}$ in a compact subset $X$ of $\mathbb{R}^n$.[5] and denote $\mathcal{X}$ the $\sigma-$algebra of Lebesgue measurable subsets of $X$. We refer to $\mathbf{C}$ as the *capital point*.[6] Denote $\mathscr{D}_X$ to be the space of positive bounded measures on $\mathcal{X}$, endowed with the topology of weak convergence, and $\mathscr{P}_X$ to be its subspace of probability measures (i.e. $\mathscr{P}_X = \{p : \int_X dp = 1\}$). Notice that $\mathscr{P}_X$ is an affine space.

We will use the term *normalized distribution*, or simply *distribution*, for all measures $p \in \mathscr{P}_X$. A normalized distribution $p_{\mathbf{x}}$ is called $\epsilon$-simple, or just simple, at a point $\mathbf{x} \in S$ if $p_{\mathbf{x}}$ is a uniform measure on the ball $\mathbf{B}(\mathbf{x}, \epsilon)$ of center $\mathbf{x}$ and radius $\epsilon$ (with $\mathbf{B}(\mathbf{x}, \epsilon) = \{\mathbf{z} : |\mathbf{z} - \mathbf{x}| < \epsilon\}$). We focus our attention on distributions whose size is normalized to one because, generally speaking, we want to be able to disentangle features of the distribution that are distinct from concentration *per se*. Most importantly among these features, of course, is the size of the population under consideration.[7]

We can now define our main object of interest, which is an order that compares distributions in relation to the capital point $\mathbf{C}$.

**Definition 1 (Concentration Order)** *A concentration order of distributions based on the capital $\mathbf{C}$ is a complete preference relation $\succsim_{\mathbf{C}}$ on $\mathscr{P}_X$, that is, a transitive binary relation on $\mathscr{P}_X$ such that for all $p$ and $q \in \mathscr{P}_X$ either $p \succsim_{\mathbf{C}} q$ or $q \succsim_{\mathbf{C}} p$.*

---

[5]While our presentation focuses on Euclidean spaces, our approach is, for the most part, easily generalizable to any normed vector spaces. (We will explicitly note the parts that are not.) We also conjecture that the approach is generalizable to a large set of "well-behaved" metric spaces. (See the discussion on normed vector spaces and metric spaces in the Appendix.) This means that we can deal with non-Euclidean metrics such as, say, the time (or cost) to travel between any two points on the map. This is a metric that satisfies the basic metric properties (identity of indiscernibles, symmetry and triangular inequality), while being complicated by road conditions, natural and institutional barriers, etc. We can thus extend our analysis to an even more general framework, which could be tailored to specific applications.

[6]We use the term "capital point", and not "center", to emphasize that said point need not be located at any spatial concept of a center, such as a baricenter or the center of a circle.

[7]We will see that it is always possible to normalize the distribution so that it is re-scaled to a unit size, and we can focus without loss of generality on such normalized distributions. This is what we will do in the remainder of the paper. Certain axiomatic frameworks, such as the construction of inequality measures, may choose to incorporate a population invariance axiom imposing the equivalence of distributions that differ only in size. We decide to leave that choice to specific applications.

We will read $p \succ_{\mathbf{C}} q$ as "$p$ is more concentrated around $\mathbf{C}$ than $q$". Our entire approach can be understood as giving further content to this relation.

In order to think about the concept of spatial concentration, it will be helpful to define some transformations on the space $\mathscr{D}_X$, namely "squeeze" (homothety) and translation. Roughly speaking, a *squeeze* brings each point in $\mathbb{R}^n$ closer (when the scaling ratio is positive and less than one) to a center point by the same proportion. A *translation* moves each point in $\mathbb{R}^n$ away by the same vector. These concepts are easily extendable to distributions, and we can define them more formally as follows:

**Definition 2 (Squeeze, or homothetic transformation)** *A squeeze of origin $\mathbf{O} \in \mathbb{R}^n$ and ratio $\rho \in \mathbb{R}$, denoted $\mathcal{S}_{(\mathbf{O},\rho)}$, is a self-map on $\mathbb{R}^n$ that brings any point $\mathbf{x}$ closer to $\mathbf{O}$ by a factor of $\rho$:*

$$\mathcal{S}_{(\mathbf{O},\rho)}(\mathbf{x}) = \rho(\mathbf{x} - \mathbf{O}) + \mathbf{O}.$$

**Definition 3 (Translation)** *A translation of vector $\mathbf{t}$, denoted $\mathcal{T}_{\mathbf{t}}$, is a self-map on $\mathbb{R}^n$ such that:*

$$\mathcal{T}_{\mathbf{t}}(\mathbf{x}) = \mathbf{x} + \mathbf{t},$$

**Definition 4 (Extension to sets and distributions)** *The squeeze and the translation are defined on the $\sigma-$algebra $\mathcal{X}$ such that for each set $Z \in \mathcal{X}$:*

$$\mathcal{S}_{(\mathbf{O},\rho)}(Z) = \{\mathcal{S}_{(\mathbf{O},\rho)}(\mathbf{z}) : \mathbf{z} \in Z\}; \quad \mathcal{T}_{\mathbf{t}}(Z) = \{\mathcal{T}_{\mathbf{t}}(\mathbf{z}) : \mathbf{z} \in Z\}.$$

*The squeeze and the translation are defined on $\mathscr{D}_X$ such that for each distribution $p \in \mathscr{D}_X$ and each set $Z \in \mathcal{X}$:*

$$\mathcal{S}_{(\mathbf{O},\rho)}(p)(Z) = p(\mathcal{S}_{(\mathbf{O},\rho)}(p)(Z)); \quad \mathcal{T}_{\mathbf{t}}(p)(Z) = p(\mathcal{T}_{\mathbf{t}}(Z)).$$

A couple of additional definitions will be helpful for our statements and proofs. First, let us define a *univariate distribution* that is associated with any $p \in \mathscr{P}_X$ and $\mathbf{C} \in X$. This univariate distribution, intuitively speaking, characterizes the probability of a point being within a certain distance from $\mathbf{C}$, under the distribution $p$. (Note that we thus define univariate distributions on $\mathbb{R}^+$, so that they should not be confused with the distributions $p \in \mathscr{P}_X$.) More formally, it is defined as follows:

**Definition 5 (Univariate distribution)** *Each $p \in \mathscr{P}_X$ is associated with a unique univariate distribution $F_p^{\mathbf{C}}$ on $\mathbb{R}^+$ characterized by the following cumulative distribution function:*

$$F_p^{\mathbf{C}}(d) = p(\{\mathbf{x} : |\mathbf{x} - \mathbf{C}| \leq d\}).$$

Note that the definition is sensible because for every $p \in \mathscr{P}_X$, the function $F_p^{\mathbf{C}}$ on $\mathbb{R}^+$ is (weakly) increasing and continuous on the right.

Our final definition is that of a *uniform ball distribution*, which is simply a uniform distribution defined over a ball.

**Definition 6 (Uniform ball distribution)** *A uniform ball distribution $p_{(\mathbf{T}, \kappa)}$ of center $\mathbf{T}$ and radius $\kappa$ is a uniform probability distribution on the ball $\mathbf{B}(\mathbf{T}, \kappa)$.*

# 3 The Expected Influence Approach

We want to give content to the concentration order $\succsim_{\mathbf{C}}$, so that it makes sense to compare distributions with regard to how concentrated they are around the capital $\mathbf{C}$. For that, it is useful to start by laying out our approach in an informal way. Since we are interested in normalized distributions, represented by the subspace $\mathscr{P}_X$, it is convenient to think of them as probability distributions. To fix ideas, let us think about the application within which we will exemplify the empirical implementation of the approach: the population of a country and its concentration around the capital city. We can thus conceptualize the distribution as describing the probability that a given individual (behind a "veil of ignorance") will end up located at any given point in the country. Our basic intuition is to think about the concentration order $\succsim_{\mathbf{C}}$ as ranking the aggregate influence that is exerted by the capital $\mathbf{C}$ on all individuals, or symmetrically the influence exerted by all points on the capital, under each distribution. This probabilistic interpretation suggests that we can resort to expected utility theory in order to think of desirable properties that such concentration order ought to display. This is the guiding principle of our *Expected Influence* approach.

The analogy with von Neumann-Morgenstern expected utility theory suggests our basic axiom, which (in consolidated form) closely parallels the standard *independence* and *continuity* axioms:

**Axiom 1 (Independence / Subgroup Consistency and Continuity)** *1. **Independence / Subgroup Consistency:***

$$p \succsim_{\mathbf{C}} q \Rightarrow \lambda p + (1 - \lambda)r \succsim_{\mathbf{C}} \lambda q + (1 - \lambda)r \quad \forall p, q, r \in \mathscr{P}_X \,, \lambda \in [0, 1].$$

*2. **Continuity:** $\forall p \in \mathscr{P}_X$, $\{q \in \mathscr{P}_X : p \succsim_{\mathbf{C}} q\}$ and $\{q \in \mathscr{P}_X : q \succsim_{\mathbf{C}} p\}$ are both closed in the topology of weak convergence.*

The intuition for this axiom is straightforward in light of the aforementioned probabilistic interpretation. In words, the independence component of the axiom states that if a distribution $p$ is more concentrated around $\mathbf{C}$ than $q$, then a distribution that combines $p$ and another distribution $r$ will be more concentrated around $\mathbf{C}$ than a distribution that combines $q$ and $r$. This is the familiar notion of independence of irrelevant alternatives: if an individual ends up located at a certain point, the points were she could have ended up, but did not, should not matter for the assessment of influence. By the same token, the continuity component means that small changes in the "probabilities" of being located in different points should not affect the concentration ranking.

The alternative label of subgroup consistency illustrates an alternative (and equivalent) interpretation for the independence axiom. Subgroup consistency can be understood again in reference to our application: suppose that we divide the population of the country into subgroups, and that one of these subgroups becomes more concentrated around the capital, while the distribution of all other subgroups remains unaffected. Subgroup consistency requires that the total population of the country be judged to be more concentrated as a result of that change. That is to say, for instance, that if the Flemish population becomes more concentrated around Brussels, and the Walloon population stays put, the population of Belgium as a whole will have become more concentrated around its capital. This is a standard property often imposed by the literature concerned with the measurement of inequality (Foster and Sen 1997) and poverty (Foster and Shorrocks 1991), and seems to be a natural starting point for our concept of concentration. While it does impose restrictions in terms of possible interactions between the concentrations of different subgroups – it essentially imposes that they are independent – this agnostic position with respect to the direction of these possible interactions seems to be appropriate for our general approach.

Axiom 1 thus states the first properties that will give content to the concentration relation $\succsim_{\mathbf{C}}$. We can now use it, with the arsenal of expected utility theory, to establish our first main result, the backbone of our Expected Influence approach:

**Theorem 1 (Expected Influence)** *The concentration order $\succsim_{\mathbf{C}}$ satisfies Axiom 1 if and only if there exists a function $h_{\mathbf{C}} : X \to \mathbb{R}$ that is bounded, continuous, and such that:*

$$p \succsim_{\mathbf{C}} q \Leftrightarrow \mathbf{I}_{\mathbf{C}}(p) \geq \mathbf{I}_{\mathbf{C}}(q) \quad \forall p, q \in \mathscr{P}_X,$$

*where $\mathbf{I}_{\mathbf{C}}$ is a real-valued function on $\mathscr{P}_X$ such that:*

$$\mathbf{I}_{\mathbf{C}}(p) = \int_X h_{\mathbf{C}} dp \quad \forall p \in \mathscr{P}_X.$$

*The function $h_{\mathbf{C}}$ is called the influence function associated with $\mathbf{I}_{\mathbf{C}}(\cdot)$. Moreover, $\mathbf{I}_{\mathbf{C}}$ and its associated influence function are unique up to positive affine transformations.*

**Proof.** From Axiom 1, the result follows from standard arguments from expected utility theory – see for instance Theorem 3.2.2 from Karni and Schmeidler (1991). ∎

This Expected Influence Theorem means that the concentration order $\succsim_{\mathbf{C}}$ can be represented by an index $\mathbf{I}_{\mathbf{C}}$ defined on $\mathscr{P}_X$ (uniquely up to an affine transformation) and its associated influence function $h_{\mathbf{C}}$ defined on $X$. The analogy with expected utility is once again instructive, as the influence function plays a role analogous to that of the (Bernoulli) utility function: it is a cardinal measure of the influence exerted by the capital on any given point.

The theorem gives us a very natural way to frame the concept of concentration around a point of interest: a distribution will be more concentrated than another when the expected influence of the capital over an individual (or vice-versa) is greater in the former than in the latter. With that in mind, we will refer to this index $\mathbf{I}_{\mathbf{C}}$ as an Expected Influence (EI) representation of the concentration relation. (Of course, every monotonic transformation of $\mathbf{I}_{\mathbf{C}}$ will still represent the same order $\succsim_{\mathbf{C}}$, but without the EI representation.)[8]

The EI representation displays a property that proves very convenient in applications, which we call decomposability and state as follows:

---

[8]It is important to keep in mind that Axiom 1 may rule out what could be interesting cases in specific applications. Just as with expected utility, it leaves open the possibility of "paradoxes" that violate independence – one could think, for instance, of a situation where the influence of an individual is disproportionately increased by the presence of other individuals in the same location. In any case, as with expected utility, Axiom 1 provides us with a general language to understand the concept, and a benchmark against which to interpret departures in specific cases.

**Corollary 1.1 (Decomposability)**

$$\mathbf{I}\left(\lambda p + (1-\lambda)q, \mathbf{C}\right) = \lambda \mathbf{I}(p, \mathbf{C}) + (1-\lambda)\mathbf{I}(q, \mathbf{C}) \quad \forall p, q \in \mathscr{P}_x \ , \ \lambda \in [0,1].$$

Decomposability means that the EI index can be computed separately for any given set of subgroups, and we can add these indices to obtain the overall measure for the entire population. This means, for instance, that the concentration of the US population around Washington, DC can be decomposed into the concentration of the population of each state around that capital point, or into the concentration of groups defined along ethnic lines, income, or any other arbitrary criterion – we can compute the index separately for each group, and from those be able to obtain the overall index for the entire population. This property is closely related to subgroup consistency, as noted by Foster and Shorrocks (1991), but it builds in more structure, with the additive feature. In particular, while any monotonic transformation of the EI index will represent the same concentration order (and hence satisfy Axiom 1), only the EI representation will be decomposable.

# 4 The Centered Index of Spatial Concentration

Having obtained the EI representation, we can now specify additional basic properties that the order $\succsim_{\mathbf{C}}$ should satisfy so that it captures the idea of concentration around the capital point $\mathbf{C}$. We will show that these basic properties impose substantial constraints on the EI indices that would be acceptable. These constraints will in turn define our class of Centered Indices of Spatial Concentration (CISC).

The first of these properties is *monotonicity*, which should be satisfied by any reasonable concept of concentration around a capital point. Suppose we compare two distributions, $p$ and $q$, and for any $d > 0$ there is more population that is within a distance $d$ from $\mathbf{C}$ under the former than under the latter. It should naturally be the case that $p$ is judged to be more concentrated than $q$. Note that this can be stated in terms of first-order stochastic dominance (FOSD): if the univariate distribution $F_q^{\mathbf{C}}$ FOSDs the univariate distribution $F_p^{\mathbf{C}}$, then the concentration order should rank $p$ ahead of $q$. This enables us to capture this idea concisely as follows:

**Axiom 2 (Monotonicity or First-Order Spatial Dominance)**

$$\exists \epsilon > 0 : \forall p, q \in \mathscr{P}_{\mathbf{B}(\mathbf{C}, \epsilon)}, F_q^{\mathbf{C}} \succ_{FOSD} F_p^{\mathbf{C}} \Rightarrow p \succ_{\mathbf{C}} q$$

10

The axiom states that there exists a neighborhood around the capital point within which FOSD implies a specific ordering of the associated distributions. Note that in this particular definition monotonicity is defined locally, that is to say in a neighborhood of the capital. This is a weaker notion than one where stochastic dominance is defined globally, and leaves open the possibility that some points would be more influential than others that are closer to the capital.

The second basic property, which we label *rank invariance*, can be understood with reference to what happens when we change the units in which distances are measured. Suppose we have two distributions, $p$ and $q$, and the distances between points are measured in miles. If $p$ is deemed to be more concentrated around $\mathbf{C}$ than $q$, it stands to reason that this relative ranking should not change if distances were instead measured in kilometers. In other words, changing the unit of distance measure should not change the ordering of distributions by the concentration order $\succsim_{\mathbf{C}}$.

This change in units is isomorphic to a squeeze of the distribution around the capital point of interest by a factor of $\rho$, where $\rho > 0$ gives us the conversion rate between the different units. (Obviously, a "squeeze" in which $\rho > 1$ is actually an expansion around the capital point.) As a result, our property affirms that the relative order of different distributions remains unchanged when they are squeezed or expanded around the capital. We state this in the following axiom:

**Axiom 3 (Rank Invariance)**

$$p \succsim q \Leftrightarrow \mathcal{S}_{(\mathbf{C},\rho)}(p) \succsim \mathcal{S}_{(\mathbf{C},\rho)}(q) \quad \forall p, q \in \mathscr{P}_X, \rho > 0.$$

This axiom embodies a property of neutrality with respect to labels: to build on the example of a Euclidean space, it requires that relabeling all the axes proportionally will leave the order unaffected. This means that we attribute independent meaning to distances, regardless of the unit in which they are measured.[9]

As it turns out, these two very natural axioms impose remarkably powerful restrictions on the set of influence functions that can be used to represent the concentration order within the EI framework. This can be stated in the following theorem, which characterizes the admissible CISCs:

---

[9]This feature is obviously natural in a geographical context, but we surmise that it is just as fundamental in abstract contexts where the specific scale is arbitrary. Consider an example in which preferences with respect to policy are measured in an arbitrary 1-5 scale. If the researcher is not ready to impose the rank invariance property were she to rescale them to a 1-10 scale, it would probably not make much sense to speak of concentration.

**Theorem 2 (CISC)** *A concentration order $\succsim_{\mathbf{C}}$ satisfies Axioms 1-3 if and only if it is represented by an EI index $\mathbf{I_C}$ with the influence function $h_{\mathbf{C}}$ such that:*

$$h_{\mathbf{C}}(\mathbf{x}) = \alpha|\mathbf{x} - \mathbf{C}|^{\gamma} + \beta \stackrel{def}{\equiv} h(|\mathbf{x} - \mathbf{C}|),$$

*where $\alpha < 0$ and $\gamma > 0$.*

*Moreover, if $\gamma < 0$, or if $h(|\mathbf{x} - \mathbf{C}|) = \alpha \log(|\mathbf{x} - \mathbf{C}|) + \beta$, and $\alpha < 0$, then the corresponding $\succsim_{\mathbf{C}}$ satisfies Axioms 2 and 3, and Axiom 1.1 (Independence); but not Axiom 1.2 (Continuity).*

**Proof.** See Appendix. ∎

Let us start by focusing on the first part of the theorem. In essence, Axioms 2 and 3 imply that the influence function that was defined in Theorem 1 must actually be a *monotonically decreasing, isoelastic function of the distance to the capital point* $\mathbf{C}$.[10] (We henceforth denote distances by $z$, for conciseness.) By isoelastic, we mean that, if we define $R_h(z) \stackrel{def}{\equiv} -\frac{h''(z)z}{h'(z)}$ as the elasticity of the marginal influence function with respect to distance – or alternatively, the "coefficient of relative risk aversion" of the influence function – then Theorem 2 establishes that our class of admissible CISCs must have $R_h(z) = 1 - \gamma \equiv R_h$, a constant.[11]

The isoelastic property is directly related to Axiom 3, i.e. rank invariance. The monotonicity property is in turn tightly linked to Axiom 2. We should note, however, that the latter axiom is local, in the sense that it refers to a neighborhood of the capital point, but the property of monotonicity is global. This is due to the combination of the two axioms, since the isoelastic shape means that local monotonicity implies global monotonicity.

**The problem of unboundedness**   The first part of Theorem 2 imposes a constraint on the elasticity parameter: $\gamma > 0$ ($R_h < 1$). This merits additional scrutiny. Theorem 1 establishes that the EI representation implies bounded influence functions, whereas functions with $R_h \geq 1$ are unbounded at $z = 0$. It is well-known from expected utility theory that unboundedness implies that expected utility (and hence expected influence, in our context) could be infinite for certain probability distributions, which necessarily violates Axiom 1.2 (Continuity).[12] This is exactly what the second part of Theorem 2 states.

---

[10]We denote this function by $h$, as distinct from $h_{\mathbf{C}}$, in order to emphasize the distinction between the two objects. However, with a slight abuse of terminology, we still refer to it as the influence function.

[11]Technically, Theorem 2 also affirms the infinite differentiability of the impact function $h$, so the expression of $R_h(z)$ is meaningful.

[12]See Kreps (1988, ch. 5) for an intuitive discussion on the difficulty of having unbounded utility functions.

As we will see in the next section, the elasticity parameter has very meaningful interpretations, and we do not want to rule out a range of possible values for that parameter on technical grounds. Fortunately, expected utility theory provides several answers to the unboundedness problem (see for instance Ledyard 1971, Fishburn 1975, 1976) that allow us to keep (at least some of) the influence functions with $R_h \geq 1$ within the set of admissible solutions. These solutions typically involve relaxing Axiom 1's Continuity, and imposing certain restrictions on the class of distributions $\mathscr{P}_X$ that are admissible. These adaptations, which we discuss in the Appendix, are very mild and, we surmise, innocuous in the vast majority of conceivable applications of the index.

In practice, any discrete dataset – which is likely to encompass the vast majority of instances of empirical implementation of the index – will allow the full set of elasticity parameter values. The one implementation requirement is that no observations be exactly at the capital point **C**, although we could have observations that are arbitrarily close to it. (We will get back to this issue when presenting our empirical implementation.) Even in continuous cases, a simple weakening of Axiom 1's Continuity allows for a wide range if functions with $R_h \geq 1$ to be admissible.

## 5   Discussion

Our three basic axioms define a circumscribed class of admissible CISCs, but before moving on to the implementation stage a few issues must be tackled. First, a crucial feature of the CISC defined in Theorem 2 is its flexibility: the degrees of freedom afforded by the parameters $\gamma$, $\alpha$ and $\beta$ mean that in any application it will be possible to shape the index in order to make it most suitable to the specific goals of the analysis. We must thus discuss how to pick those parameter values, and for that we first need an *interpretation for the elasticity* parameter $\gamma$. We also want to have a framework within which to think of the free parameters $\alpha$ and $\beta$, which we will do in the context of the *normalization of the index*. In addition, we will also provide a *comparison of the CISC with alternative measures* of concentration and with other related concepts.

## 5.1  Interpreting the Elasticity Parameter

### 5.1.1  Elasticity and Marginal Influence

The first thing to note about the elasticity parameter $\gamma$ is that it has to do with how the marginal influence function is affected by the distance to the capital. To make things more concrete, suppose we consider bringing an individual closer to the capital along the ray going through her initial location, and ask what is the impact of this movement on expected influence. The value of $\gamma$ determines how this impact is affected when we vary the initial point of that movement: if we move an individual 10 miles closer to the capital, how does it matter whether she started 20 or 200 miles away?

As it turns out, our framework enables us to understand this choice in terms of the marginal influence at the capital $\mathbf{C}$ relative to other points in $X$. Two very mild conditions on that relative marginal influence will allow us to partition the parameter space into one region in which the answer to the question above would be that the impact is greater when the movement starts closer to the capital, and another region in which the answer would be the opposite.

**Condition 1 (Maximal Marginal Influence)** $\exists \mathbf{x} \neq \mathbf{C}$, $\eta > 0$, and $\mathbf{t} = \eta(\mathbf{C} - \mathbf{x})$ such that:

$$\frac{1}{2}p_{\mathbf{C}} + \frac{1}{2}p_{\mathbf{x}} \succsim_{\mathbf{C}} \frac{1}{2}p_{\mathcal{T}_{\mathbf{t}}(\mathbf{C})} + \frac{1}{2}p_{\mathcal{T}_{-\mathbf{t}}(\mathbf{x})}$$

**Condition 2 (Minimal Marginal Influence)** $\exists \mathbf{x} \neq \mathbf{C}$, $\eta > 0$, and $\mathbf{t} = \eta(\mathbf{C} - \mathbf{x})$ such that:

$$\frac{1}{2}p_{\mathbf{C}} + \frac{1}{2}p_{\mathbf{x}} \precsim_{\mathbf{C}} \frac{1}{2}p_{\mathcal{T}_{\mathbf{t}}(\mathbf{C})} + \frac{1}{2}p_{\mathcal{T}_{-\mathbf{t}}(\mathbf{x})}$$

The first condition states, in a nutshell, that there exists a point at which the marginal influence is (weakly) smaller than at the capital. In ordinal terms, we consider a distribution that consists of a mass point at $\mathbf{C}$ and another mass point at some other point $\mathbf{x}$. If we slightly shift the former mass point away from $\mathbf{C}$, while shifting the latter towards $\mathbf{C}$ by the same distance and in the same direction, the condition implies that the resulting distribution will be less concentrated around $\mathbf{C}$ than the original one. This is what underlies the idea that the marginal influence at $\mathbf{x}$ is smaller than at $\mathbf{C}$. The second condition states that there exists a point at which that marginal influence is (weakly) greater than at the capital, i.e. the same shift will result in a more concentrated distribution. These two conditions, plus Axioms 1-3, yield the following:

**Proposition 1 (Convexity / Concavity)** *Suppose Axioms 1-3 hold, then:*

1. *Condition 1 holds if and only if $R_h \geq 0$ (or equivalently, $\gamma \leq 1$).*

2. *Condition 2 holds if and only if $R_h \leq 0$ (or equivalently, $\gamma \geq 1$).*

**Proof.** See Appendix. ∎

Condition 1 thus implies that the influence function will be a convex function of the distance to the capital, whereas Condition 2 implies the concavity of that function.[13] The former case means, intuitively, that movements that occur close to the capital point will have a (weakly) greater weight than similar movements occurring farther away. The latter case has the opposite implication, and it is obvious that the two can only happen simultaneously in the case where $R_h = 0$ ($\gamma = 1$). We will refer to this limit case, in which all points display the same marginal influence, as the *Linear* CISC (L-CISC). We state it here for future reference:

**Definition 7 (Linear CISC)** *The Linear Centered Index of Spatial Concentration (L-CISC) is the special case of the CISC defined in Theorem 2 in which $\gamma = 1$. The influence function is defined as:*

$$h(z) = \alpha z + \beta,$$

*with $\alpha < 0$.*

A brief inspection shows that Conditions 1 and 2 are mild enough that, in principle, they could clearly hold simultaneously even beyond the boundary case. It is only when combined with our basic axioms (particularly Axiom 3) that this ceases to be possible. In other words, with the exception of the boundary case that yields the L-CISC, having both conditions hold requires giving up rank invariance, and accepting the possibility that changes in scale will affect the ranking of distributions.

In sum, Proposition 1 means that $R_h$ or $\gamma$ are sufficient statistics when it comes to describing the relative weight that different points will have in affecting the ranking of distributions, depending on their distance to the capital. Different contexts might call for different values,

---

[13]The convex case generated by Condition 1 includes influence functions that can run into the unboundedness issue highlighted in the previous section. More specifically, this is the case for $R_h \geq 1$ ($\gamma \leq 0$). As pointed out in the previous discussion, using these indices will require imposing some restrictions on the admissible distributions to avoid unboundedness. Also as previously discussed, the discrete-support distributions that are likely to arise in most instances of empirical implementation are always admissible, so that any elasticity parameter can be used.

but we know that higher $R_h$ implies attaching a greater marginal influence to closer points, and that negative values for $R_h$ mean that more distant points have greater marginal influence.

### 5.1.2 Elasticity and Mean-Preserving Spreads

We can gain further intuition about the elasticity parameter by extending the analogy with expected utility theory to include the concept of second-order stochastic dominance (SOSD), just as we have done with FOSD. We know from the study of choice under uncertainty that there is a close connection between the curvature of utility functions and the ranking of mean-preserving spreads (MPS) of distributions, as captured by the concept of SOSD. We can now show that extending this concept to our setting yields another interpretation for the elasticity.

The first challenge in to extend the concepts of MPS and SOSD for general, $n$-dimensional Euclidean spaces, which is crucial since our applications will often involve more than one dimension.[14] We must thus extend the concepts to that multidimensional context, which is not obvious because it is not immediate to conceptualize a MPS in many dimensions.

A natural way would be to use the concept of univariate distributions associated with the distributions $p \in \mathscr{P}_X$, and directly define SOSD over them. It will turn out to be more fruitful to start instead from what we may call a "uniform MPS", which we define in terms of a squeeze of the uniform ball distribution that we have defined. Informally speaking, consider individual observations that are uniformly distributed over a ball around a given point $\mathbf{T}$, such that $\mathbf{C}$ is not in that ball. Now suppose we squeeze this distribution around that point, such that the resulting distribution is also uniform over the resulting ball.[15] (Of course, once again this could be a squeeze or an expansion ("spread") depending on the coefficient associated with the transformation.)

The interesting question is how the concentration order $\succsim_{\mathbf{C}}$ will rank the distribution that results from this uniform MPS. Let us consider the case where we have $\rho < 1$ (i.e. a squeeze): should the resulting distribution be ranked as more or less concentrated *around* $\mathbf{C}$ than the initial distribution? The answer to this question must be context-specific – still pushing the analogy with choice under uncertainty, we could be in a case of "risk averse" or "risk loving" behavior. But we can show that it is intimately related to the elasticity parameter, in a way

---

[14]This is in fact the only part of our approach that is restricted to Euclidean spaces.

[15]Other types of MPS could be envisioned, in ways that would not preserve uniformity. We will see shortly why this is a convenient definition.

that enables us to uncover some additional interesting properties.

As it turns out, the answer relies on two conditions that have a similar flavor to Conditions 1 and 2. Their implications can be studied with reference to harmonic function theory, which is why we label then sub- and super-harmonicity. They are as follows:[16]

**Condition 3 (Sub-Harmonicity)**

$$\mathcal{S}_{(\mathbf{T},\rho)}(p_{(\mathbf{T},\kappa)}) \precsim_{\mathbf{C}} p_{(\mathbf{T},\kappa)} \ \forall \rho < 1, \mathbf{B}(\mathbf{T},\kappa) \not\ni \mathbf{C}.$$

**Condition 4 (Super-Harmonicity)**

$$\mathcal{S}_{(\mathbf{T},\rho)}(p_{(\mathbf{T},\kappa)}) \succsim_{\mathbf{C}} p_{(\mathbf{T},\kappa)} \ \forall \rho < 1, \mathbf{B}(\mathbf{T},\kappa) \not\ni \mathbf{C}.$$

Condition 3 states that the distribution that results from a uniform squeeze around $\mathbf{T}$ is (weakly) less concentrated around $\mathbf{C}$ than the original distribution; Condition 4 states the opposite. Note that Condition 4 is akin to SOSD: any (uniform) MPS will lead to a distribution that is ranked below ("dominated by") the original distribution. Condition 3 inverts that. This in turn suggests that these conditions will be related to the curvature of the influence function, just as SOSD is linked to the concavity of utility functions. This is indeed the case, but the multidimensional uniform extension leads us into a result that is subtly but importantly different:[17]

**Proposition 2 (SOSD)** *Suppose Axioms 1-3 hold, then:*

1. *Condition 3 holds if and only if $R_h \geq n - 1$ (or equivalently, $\gamma \leq 2 - n$).*

2. *Condition 4 holds if and only if $R_h \leq n - 1$ (or equivalently, $\gamma \geq 2 - n$).*

**Proof.** See Appendix. ∎

This proposition thus provides us with another interpretation for the elasticity parameter, having to do with MPS, in addition to the interpretation focusing on marginal influence.[18] As

---

[16]Note that the definitions explicitly restrict attention to uniform MPS defined for a ball around $\mathbf{T}$ that does *not* contain $\mathbf{C}$. We will return to this point shortly.

[17]Another way to express the Proposition is to say that Condition 3 implies that the influence function is sub-harmonic, whereas Condition 4 implies that the influence function is super-harmonic. See Appendix.

[18]Note in particular that there are functions that are convex, but super-harmonic; in other words, they represent concentration orders under which a uniform MPS results in a function that is ranked above ("dominates") the original function.

17

before, the specific choice for the parameter can be thought of as context-specific.[19] but it turns out that the "harmonic" case of $R_h = n - 1$ displays interesting properties. Intuitively speaking, it is a case in which any local force of concentration at an arbitrary point $\mathbf{T}$ does not affect the ranking of distributions in terms of their concentration around $\mathbf{C}$. In that sense, the index measures the "gravitational pull" exerted by the capital point, while disentangling from the data any impact of the presence of other local forces. It is invariant to changes in the degree of attraction, or "gravity", of any other points. An analogy comes from the gravitational pull of the Sun over the Earth, which is measured focusing only on the distance between the two bodies and their characteristics (mass), leaving aside the influence of all other planets. For this reason, we will refer to this special case as the *Gravity-based* CISC (G-CISC).[20]

Besides the helpful interpretation, the G-CISC has a convenient property in terms of the implementation of the index. In practice, the information used to compute the index typically comes in a grid format, where we only know the aggregate information of each cell. This introduces a source of measurement error. The G-CISC is orthogonal to symmetric measurement error, because it satisfies both Condition 3 and 4 – in other words, the influence function is harmonic.

**Definition 8 (Gravity-based CISC)** *The Gravity-based Centered Index of Spatial Concentration (G-CISC) is the special case of the CISC defined in Theorem 2 in which $\gamma = 2 - n$. The influence function is defined as:* $h(z) = \begin{cases} \alpha z + \beta \ \forall z > 0, \ \alpha < 0 & \text{if } n = 1 \\ \alpha \log z + \beta \ \forall z > 0, \ \alpha < 0 & \text{if } n = 2 \\ \alpha z^{2-n} + \beta \ \forall z > 0, \ \alpha > 0 & \text{if } n \geq 3 \end{cases}$.

---

[19]One caveat may be in order, however. If we look at uniform MPS defined for a ball around $\mathbf{T}$ that does contain $\mathbf{C}$ – which were not considered by Conditions 3 and 4 – it turns out that it is not possible to pin down what will happen to the index when the influence function is sub-harmonic. This is because the monotonicity with respect to point $\mathbf{C}$ (implied by Axioms 2 and 3) pushes in the opposite direction of the sub-harmonicity. This indeterminacy could be thought of as an argument for focusing on the super-harmonic case ($R_h \leq 1$).

[20]The analogy with gravity is actually deeper than what might seem at first sight: there is a connection between the G-CISC and the concept of *potential* in physics, which refers to the potential energy stored within a physical system – e.g. gravitational potential is the stored energy that results in forces that could move objects in space. In fact, our G-CISC can be interpreted, roughly speaking, as a measure of the potential associated with the capital point of interest. This is because the G-CISC is the index that satisfies the Laplace equation, and potential is defined by a solution to that equation – those readers familiar with the physics of the concept will have noticed the connection from the use of harmonic function theory. To see this connection more clearly, note that in the "real life" three-dimensional space of Newtonian mechanics, the potential of a point with respect to a mass is the sum of inverse distances from that point to each location in the mass – and since the gravity force is the derivative of potential, it is proportional to the inverse of the squared distance, as one may recall from the classical Newtonian representation. Our G-CISC in three-dimensional space is also the weighted sum of the inverse of distance, and thus coincides with the concept of potential in physics.

## 5.2 Normalization

It is useful to normalize the index so that it can be easier to interpret, and this normalization should be suitably chosen in order to address the questions that are relevant in the specific context. The EI approach affords us two sets of levers for this normalization. First, there is the focus on normalized distributions, i.e. probability measures. Second, we have the degrees of freedom afforded by the CISC as defined in Theorem 2, in addition to the elasticity parameter, which are the choice of parameters $\alpha$ and $\beta$.

In order to see both angles more clearly, and motivated by our empirical implementation, let us fix ideas by focusing on a situation where our index is applied to the concentration of the population of a given geographic unit of analysis (e.g. a country) around a point of interest (e.g. the capital city), in two dimensions. Furthermore, we start by focusing our interest on the special case of the G-CISC – which means that the influence function is given by $h(z) = \alpha \log(z) + \beta$. (It is straightforward to extend the following analysis to other contexts.)

In this context, the first thing to note is that there is often other information that we may want to disentangle from population concentration *per se*, e.g. population size and the geographical size of the country. This is exactly why our theory focuses on normalized distributions, i.e. probability measures. In practice, any implementation requires that the actual distribution be normalized so that it can be expressed as a probability distribution, containing only the information we are interested in. This will be context-specific, since each context will determine what kind of information we want to leave out.

Second, we would like to have an easily interpretable scale, and this is what can guide the choice of $\alpha$ and $\beta$. A convenient rule would be to restrict the index to the $[0, 1]$ interval, with 0 and 1 representing situations of minimum and maximum concentration, respectively. The latter can be defined simply as a situation in which the entire population is located arbitrarily close to the capital, but the former will typically vary greatly with context. One could think of it as a case in which the entire population is located as far from the center as possible, but "as far as possible" could have different meanings, as we will discuss below. Alternatively, one could think of minimum concentration as a situation where the population is uniformly distributed over the entire country, along the lines of Ellison and Glaeser's (1997) "dartboard" approach.[21] This choice of parameters will inevitably be context-specific too.

---

[21]An implementation of this benchmark is available upon request.

As a general principle, we can proceed with normalization in two steps: (1) normalize the distribution under analysis, transforming it into a distribution $p \in \mathscr{P}_X$ that contains only the information we are interested in; and (2) set the parameters $\alpha$ and $\beta$ to fit a $[0, 1]$ scale. The specifics of each of these steps will depend on the context, so in what follows we discuss a few benchmark examples that will be relevant for our application, and that can illustrate the general procedure.

**Maximum distance across units of analysis**   A first approach is to set the minimum concentration based on the maximum possible distance between a point and the capital city in any of the countries for which the index is to be computed. In this case, the index is evaluated at zero if the entire population lives as far away from the center as it is possible to live in any country.[22] (As a result, only one country, the one where this maximum distance is registered, could conceivably display an index equal to zero.) This will be appropriate if we want to compare each country's concentration against a single benchmark.

In order to achieve this, the two steps are:

1. Normalize the distribution by dividing it by total population size, which means that we will be taking each country to have a population of size one.

2. Set:
$$(\alpha, \beta) = \left( -\frac{1}{\log(\overline{z})}, 1 \right)$$

   where $\overline{z} \equiv \max_{x_i, i} |x_i - \mathbf{C}_i|$ is the maximum distance between a point and the center in any country. This means that we take the (logarithm of the) largest distance between a point and the capital in any country to equal to one.

**Maximum distance within unit of analysis**   Another possibility is to evaluate the index at zero for a given country if its entire population lives as far away from the capital as it is possible to be in that particular country. This is appropriate if we want to compare each country's actual concentration to what its own conceivably lowest level would be.

With that in mind, the two steps are now:

---

[22]For instance, in the cross-country implementation later in the paper, the largest recorded distance from the capital city within any country is between the Midway Islands and Washington, DC, in the United States.

1. Besides the standard normalization by population size, normalize by $\log(\overline{z}_i)$, where $\overline{z}_i \equiv \max_x |x - \mathbf{C}_i|$ is the maximum distance between a point in country $i$ and that country's capital. This means that we not only take each country's population to be of size one, but also that we take the (logarithm of the) largest distance between a point and the capital of that country to be one as well.

2. Set:
$$(\alpha, \beta) = (-1, 1).$$

Our empirical implementation will illustrate both of these cases of normalization and their different interpretations.

## 5.3 Comparison with Other Indices

### 5.3.1 Comparison with other Centered Indices

The first obvious comparison is to other centered indices of spatial concentration used in the literature. The EI approach in fact gives us a coherent framework to think of these alternative measures, and to understand their properties and limitations.

Let us start with a class of widely used measures of spatial concentration that essentially attach zero weight to observations located at more than a certain distance from the capital point. A typical application of such a measure is the share of population of a country that lives in the capital city (often refereed to as "capital primacy"), or in the main city in that country, as in Ades and Glaeser (1995), Henderson (2003), among many others. It is easy to check that this class of measures violates monotonicity (Axiom 2), and also that it violates continuity (Axiom 1.2). Intuitively speaking, this is so because this class of measures essentially discard a lot of information. Our approach, in stark contrast, incorporates all of that information, with enough flexibility to allow for different weights according to the application – as parameterized by the elasticity of the marginal impact function with respect to distance (as in Proposition 1): the greater that elasticity, the less weight is attached to observations that are farther away from the capital point.

We can also use the EI approach to make sense of other widely used measures that do not discard that type of information. One example is the negative exponential function that is widely used in the empirical urban economics literature to describe the density of population or

economic activity (e.g. Clark 1951, Mills 1970). (This has also been used in political economy, as in Busch and Reinhardt (1999).)It has been recognized that this standard density function is very restrictive, and alternative implementations have been proposed (McMillen 2004), but (to the best of our knowledge) not from a theoretically grounded approach. With our EI framework, it can be easily verified that this measure violates rank invariance (Axiom 3), and can thus lead to different rankings as a result of changes in units.

Another interesting example, as illustrated by Galster et al. (2001), uses the inverse of the sum of the distances of each observation to the center as a measure of "centrality" in the context of studying urban sprawl. *Prima facie*, it is apparent that this measure does not fit the EI representation, because of the inverse operation. A closer look reveals, however, that this measure is a monotonic transformation of the L-CISC ($R_h = 0$). As such, Theorem 1 implies that it must represent the same concentration order that is represented by the L-CISC. This means that the Galster et al. (2001) index does satisfy – or more accurately, it represents a concentration order that satisfies – all of our axioms.

These desirable properties were certainly not obvious from the *ad hoc* approach used in the construction of the index. Our axiomatic approach enables us to identify them, and guarantee that they are satisfied in very general spaces, without being limited by an intuitive grasp of the features of a specific, concrete application. Our approach also enables us to understand in greater depth the properties of such an *ad hoc* index: in the case of Galster et al. (2001), following Propositions 1 and 2, we can state that their index attaches the same marginal influence regardless of distance, and that their measured concentration would decrease as a result of a uniform mean-preserving spread.

In sum, the EI approach provides a unified "language" to codify the concept of spatial concentration around a capital point. This language has very wide applicability, and it provides us not only with a family of CISCs but also with a way of making sense of pre-existing *ad hoc* approaches.

### 5.3.2 Comparison with Non-Centered Indices of Concentration

Besides the obvious distinction that our index is built on the concept of a particular "center", it also contains considerably more spatial information than non-centered indices. For instance, let us consider the many indices of concentration that are based on measures first designed

to deal with inequality – such as the *location Gini coefficient* calculated on a distribution of cells.[23] Such measures do not take into account the actual spatial distribution.

Indeed, consider a thought experiment where half of the cells contain exactly one individual observation, and the other half contain zero observations. Such indices do not make any distinction between a situation in which the former cells are all in the East and all the latter cells are in the West, and another situation in which both types are completely mingled together in a chess-board pattern. Generally speaking, this type of measure fails to take into account the relative positions between the cells, which can be highly problematic in many circumstances.[24] The same can be said of cruder measures of concentration, such as population density.[25] In addition, these measures are also highly non-linear with respect to individuals, because they contain a function of the cell distribution. In that sense, they are not grounded at the individual level in the way that our CISC is.

These differences are highlighted by the fact that we can actually derive a non-centered index from our centered measure. We can do so by averaging the CISC over all possible capital points (i.e. all points) within the support of the distribution, a feasible task in most applications. On $\mathbb{R}$ (say, applied to individual income distributions), this aggregative, non-centered index coincides with the much familiar Gini index of inequality. For now, we leave further explorations on this subject for future work.[26] In any case it is not possible to follow the reverse path and obtain a centered index from a given non-centered measure. This underscores the versatility of our approach.

### 5.3.3   Comparison with Related Indices: Inequality, Polarization, Riskiness

Finally, it is worth noting the links between our index and other indices designed to measure other aspects of distributions, be they spatial or not. The connection with inequality measures

---

[23]For instance, the *location Gini* is used, *inter alia*, in the context of economic geography (Krugman 1991, Jensen and Kletzer 2005), studies of migration (Rogers and Sweeney 1998), political economy (Collier and Hoeffler 2004).

[24]In practice, for instance, if we measure the concentration of the US auto industry around Detroit, it matters whether car plants are in nearby Ohio or in distant Georgia. However, a non-centered index of concentration computed at the state level would stay unchanged if all the plants in Ohio were moved to Georgia and vice-versa.

[25]A related index that does use spatial information is the measure of compactness developed by Fryer and Holden (2007).

[26]We should note that to fully develop a non-centered version of our approach, we would have to evaluate the axioms from that non-centered perspective. For instance, we conjecture that such a framework should probably add super-harmonicity (Condition 4) as an axiom, since we would not want to have that uniform MPS would lead to more concentration.

has already emerged from the very fact that such measures are used to capture spatial concentration, and we have noted that a natural extension of our approach to a non-centered setting highlights an interesting connection with the Gini coefficient.

Our approach is also interestingly related to the work on polarization by Duclos, Esteban and Ray (2004). Intuitively speaking, one could think of a notion of "polarization" as broadly working in the opposite direction of "concentration". Keeping that in mind, our concentration order would satisfy their axioms 1 and 3 (both related to our Monotonicity axiom) and also their axiom 4 (related to our Rank Invariance axiom), as an "anti-polarization" order. Their axiom 2, however, would work in the opposite direction of our Monotonicity axiom, which is of course due to the fact that polarization and concentration around a point do capture different concepts. Nevertheless, the analogy with the other three axioms suggests that our approach could help build a generalization of their measure to multidimensional contexts.

The EI approach also entails interesting connections with the measurement of riskiness – as should be evident from the connection with expected utility, and as was highlighted in our discussion of the elasticity parameter. A desirable property of a measure of riskiness, as spelled out by Aumann and Serrano (2008) is that it respects first- and second-order stochastic dominance. Axiom 2 (Monotonicity) is clearly linked to FOSD – particularly if we were to state it in a global, rather than local version. Just as previously highlighted, our Conditions 3 and 4 (and hence Proposition 2) are closely related to the notion of SOSD, and could be thought of as entailing a generalization of that concept to multidimensional contexts.

# 6    Application: Population Concentration around Capital Cities

Having established the EI approach, derived the CISC, and discussed its properties, we now move on to illustrate its applicability in practice. We focus on the distribution of population around capital points of interest – capital cities across countries and across US states, and the political center (e.g. the location of city halls) in US metropolitan areas. We will discuss descriptive statistics and basic correlations with variables of interest, and also how the index can be used to shed light on competing theories, using as an example the determinants of the location of capital cities.

## 6.1 Cross-Country Implementation

In our first application, we calculate population concentration around capital cities across countries in the world. We use the database *Gridded Population of the World* (GPW), Version 3 from the Socio-Economic Data Center (SEDC) at Columbia University. This dataset, published in 2005, contains the information for the years 1990, 1995 and 2000, and is arguably the most detailed world population map available. Over the course of more than 10 years, these data are gathered from national censuses and transformed into a global grid of 2.5 arc-minute side cells (approximately 5km), with data on population for each of the cells in this grid. [27] Details on these and all other variables used in the analysis can be found in the Data Appendix.

The first step in the analysis is to decide which version of the index to focus on. As we have shown, this entails a choice of the elasticity parameter, and of the appropriate normalization. We would expect that it will oftentimes be convenient to have a specific model that might lead to a specific choice. Since our goal in this paper is to broadly illustrate the application of the index, we choose to retain generality and implement (and compare) a broader array of indices. In any case, the choice of this broader array can exemplify the general principles discussed in the previous section in their practical implementation.

With respect to the elasticity parameter, let us first consider the interpretation focused on the marginal influence as a function of distance. We are mostly interested in the capital city as a national political center. In this case, it seems natural to assume that the marginal influence would be highest in the capital. To take a concrete example, if we want to measure the concentration of Russia's population around Moscow, it stands to reason that a given movement of individuals towards the capital would not matter less if it were to happen in the outskirts of Moscow than if it were to happen in Vladivostok, in the far east corner of the country. This assumption, as we have seen, places us in the range of convex influence functions – namely with $R_h \geq 0$ ($\gamma \leq 1$).[28] Within this range, one focal point is the limit case of the L-CISC ($R_h = 0$), which would attach equal marginal weights regardless of distance. While somewhat extreme,

---

[27] We limit our analysis to countries with more than one million inhabitants, since most of the examples with extremely high levels of concentration come from small countries and islands. The results with the full sample are very much similar, however, and are available upon request.

[28] As an example of a model that could pin down the elasticity parameter in this context, one could look at the simple model of revolutions sketched in Campante and Do (2007). In that model, political influence is modeled as a cost of joining a rebelion, which depends on the distance to the political center. The standard assumption of convex costs would naturally lead to convexity, and specific functional forms would translate into different elasticity parameters.

it also has the advantage of providing a natural benchmark for comparison, since it represents the same order as the alternative measure used for instance by Galster et al. (2001).

The second interpretation can help us expand our array, still within the set of convex influence functions, to include an example in which the marginal influence at the capital is strictly greater than anywhere else. As far as how we should think about uniform MPS in terms of political influence relative to the capital, we choose to stick with the "agnostic" position regarding the direction of the impact. This leads us to the limit case of the G-CISC – which in our two-dimensional application corresponds to $R_h = 1$, the logarithmic influence function.[29] Following the argument in the previous section, this is also a natural focal point in light of the gravity interpretation and the robustness to (symmetric) measurement error.

When it comes to normalization, we use the two cases highlighted in section 5.2. The first version ($GCISC_1$) is normalized by the maximum distance across countries, and the second version ($GCISC_2$) is normalized by the maximum distance within the country. The former case captures concentration relative to what it could possibly be in any country, while the latter captures concentration relative to what it could possibly be in that specific country.[30] In the interest of brevity, for the L-CISC we will only compute the version normalized by the maximum distance across countries, $LCISC_1$.

We will compare our indices with two alternative measures of concentration. The first alternative is the location Gini coefficient ("$Gini\_Pop$"), a non-centered measure that is often used in the literature, and the second one is the share of the population living in the capital city ("$Capital\_Primacy$"), which provides a benchmark for comparison with another very simple centered index.

### 6.1.1 Descriptive Statistics

Table 1 shows the basic descriptive statistics for the different measures, for the three years in the sample, and Table 2 presents their correlation. The first remarkable fact is that there is

---

[29]Recall that the log function is in the range of parameters that might face the unboundedness problem. Simply put, in practice the influence function is not defined for the distance of zero. In accordance with the solutions that we have discussed, what we need to do is to choose the arbitrary size $\epsilon$ of the neighborhood around the capital in which we will truncate the distribution. We pick the value of 1 (kilometer). In other words, we assign the value 1 (kilometers) to the distance from each grid center to the capital whenever it is measured at less than 1, truncating the log distance function at 0: $h(z) = \alpha \log \max\{z, 1\} + \beta = \alpha \max\{\log z, 0\} + \beta$. There is very little change in the index when we replace 1 by 10.

[30]While the non-normalized measure may present some interest in itself, we do not report it because of its extremely high correlation with population size, which prevents us from disentangling any independent effect.

very little variation over that span of time: the autocorrelation is extremely high, and almost all variation comes from the cross-country dimension. This suggests that the pattern of population distribution is fairly constant within each country, and that a period of 10 years may be too short to see important changes in that pattern. For this reason, we choose to focus on one of the years; we choose 1990 because it is the one that has the highest quality of data, as judged by the SEDC.[31]

<div align="center">[<strong>TABLES 1 AND 2 HERE</strong>]</div>

Let us start by comparing the basic properties of our indices with those of the comparison measures, noting that the appropriate benchmark for comparison in the case of G-CISC is $GCISC_1$, and not $GCISC_2$, since both location Gini and $Capital\_Primacy$ do not normalize by the geographical size of each country. The striking fact that immediately jumps from Table 2 is that our index captures a very different concept from what the location Gini is capturing: they are *negatively* correlated, both in the case of G-CISC and L-CISC. This underscores the point that typical measures of concentration are ill-suited for getting at the idea of concentration around a given point.

This point becomes even more striking when we compare the list of countries with very high and very low levels of concentration, which are displayed in Table 3. We can see that the list of the countries whose population is least concentrated around their capital cities accords very well with what was to be expected: these are by-and-large countries where the capital city is not the largest city. (The exceptions are Russia, on which we will elaborate later, and the Democratic Republic of the Congo, formerly Zaire, whose capital is located on the far west corner of the country.) By the same token, the list of highly concentrated countries is quite intuitive as well, with Singapore leading the way. The same list for the location Gini, in contrast, surely helps us understand why the correlation between the two is negative. It ranks very highly countries that have big territories and unevenly distributed populations. While this concept of concentration may of course be useful for many applications, it is quite apparent that using non-centered measures of concentration can be very misleading if the application calls for a centered notion of concentration.[32]

---

[31]Our results are very similar when we use the other two years.

[32]It is also worth noting that a measure such as location Gini is quite sensitive to how "coarse" the grid that is being used to compute the index is: the fewer cells there are, the lower the location Gini will tend to be. Our index, on the other hand, has the "unbiasedness" feature that we have already discussed.

[**TABLE 3 HERE**]

In the case of the alternative centered measure of concentration, *Capital_Primacy*, Table 2 shows that the correlation is positive, though not overwhelming.[33] Table 3 shows, however, that the ranking of countries that emerges from this measure is completely different from the ones generated by both CISCs.[34] This is not surprising, in light of the amount of information that is being discarded by *Capital_Primacy*, but another crucial problem with such a coarse measure is clearly apparent from the table: its arbitrariness. Note that Kuwait, which is one of the most concentrated countries in the world according to both CISCs, shows up as one of the *least* concentrated ones as judged by *Capital_Primacy*. This is so because the population of what is officially considered as Kuwait City, the capital, is just over 30,000, while the population of the metropolitan area is over two million. This difference of two orders of magnitude is simply due to an arbitrary delimitation of what counts as the capital city. This clearly illustrates the dramatic distortions that can result from discarding relevant information.

We can also compare the two versions of our index, L-CISC and G-CISC, in order to understand the consequences of changes in the elasticity coefficient. Table 2 shows that the correlation between $GCISC_1$ and $LCISC_1$ is positive and quite high, which is reassuring since they both purport to measure the same concept. Nevertheless, there are important empirical differences between the two. The first such difference can be seen from Figure 1, which plots histograms of both indices. We can see from the figure that the distribution of $LCISC_1$ is very skewed, whereas $GCISC_1$ has a more compelling bell-shaped distribution. This implies that the latter is generally less sensitive to extreme observations. Another way to illustrate this difference is to consider a specific comparison, between Brazil and Russia. Russia's capital, Moscow, is the country's largest city, and is located at about 600km (slightly less than 400 miles) from the country's second largest city, St Petersburg. In contrast, Brazil's capital, Brasília, is now the country's sixth largest city, and is around 900km (more than 550 miles) away from the country's largest cities, São Paulo and Rio de Janeiro, whose combined metropolitan area

---

[33]Note also that the maximum value of *Capital_Primacy* is greater than one. This is due to the fact that the data for capital city population and total population, used to compute the share, come from different sources; that data point corresponds to Singapore, which should obviously be thought of as having a measure of 1.

[34]Note that here we use the measure of *Capital_Primacy* as computed for 1995. This is because there are many fewer missing values for 1995 than there are for 1990.
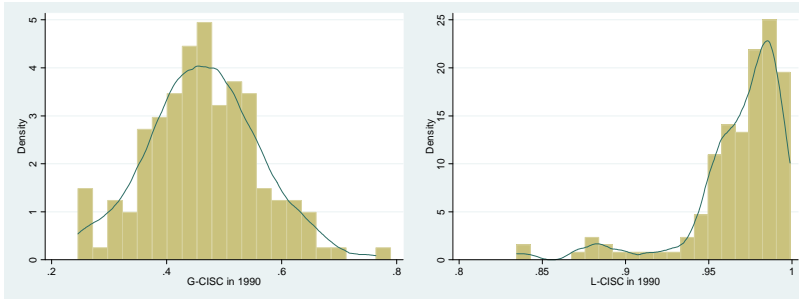
Figure 1: G-CISC and L-CISC

population is about ten times as large as Brasília's.[35] Table 3 shows that Brazil is ranked to have lower concentration than Russia with $GCISC_1$, but not with $LCISC_1$. This is because $LCISC_1$ gives a larger weight to people who are very far from the capital point of interest; roughly speaking, it gives a relatively large weight to people who are in Vladivostok. This example drives home the point that different choices of the elasticity coefficient lead to different characterizations, thus illustrating the flexibility of our approach.

Finally, we also note an interesting pattern emerging from Table 3, regarding the "size-normalized" version of our G-CISC, $GCISC_2$: the countries with the most concentrated populations seem to be fairly small ones (in terms of territory). This does not arise from "mechanical" reasons, first of all because the measure is normalized for size – the pattern suggests that the population of relatively small countries is more concentrated than that of large ones, *relative* to what it could be. In addition, while the measure for these countries may be less precise because of the small size, and consequent smaller number of grids, we know that our index is unbiased to classical measurement error. We will explore this pattern more systematically in our regression results. We can also note that $GCISC_1$ is typically much higher than $GCISC_2$: a country will have a more concentrated population relative to the maximum distance across

---

[35]According to official data, the metro area population of São Paulo, Rio de Janeiro, and Brasília is around 19 million, 12 million, and 3 million, respectively.

countries than to the maximum distance within the country itself.

### 6.1.2 Regression Analysis

We can also investigate the correlation patterns of our indices with several variables of interest. We will stop short of providing a discussion of causal inference, as it falls outside the scope of this paper, but we can nevertheless provide some interesting results that can be built upon by future research.

**Economic variables**  We start by regressing G-CISC and L-CISC on a number of economic variables of interest.[36] The results are described in Table 4. The first thing to note is that there is a *negative* correlation between land area and concentration around the capital city: countries with larger territories have populations that are less concentrated around the capital. This correlation is robust to the inclusion of a number of controls. It is also worth noting that the correlation between land area and concentration is *positive* when the latter is measured by the Gini coefficient, which is not surprising in light of Table 3, but nevertheless underscores the point that using Gini as a proxy for concentration of population around the capital city is deeply misleading.

### [TABLE 4 HERE]

It is not that surprising that the measures that are not normalized for size will indicate a negative correlation with territorial size. However, our $GCISC_2$ index, which is normalized, also displays a very significant and robust negative correlation, as anticipated from Table 3, which suggests that such correlation is more than a mechanical artifact of the construction of the indices.

The second robust correlation pattern displayed by the different versions of our CISC is as follows: there is a *negative* correlation between the size of population, and how concentrated it is around the capital. In other words, the smaller the country's population is, the more concentrated it is around the capital. One can speculate over the reasons behind this negative correlation; perhaps countries with larger populations are more likely to have other centers of attraction that lead to the equilibrium distribution of population being more dispersed around

---

[36]All of the variables that are time-variant are measured with a 5-year lag in our main specifications. Experimenting with other lags did not affect the results. All control variables are described in the Appendix.

the capital city. (We should note, however, that in the case of G-CISC the property of Gravity, which isolates the attraction exerted by the capital point of interest, ensures that the existence of other centers of attraction will not be mechanically built into the index.) It is worth noting that the relationship is weaker for $GCISC_2$, where concentration is normalized by the territorial size of the country. These patterns can and should be the subject of future research.[37]

**Governance variables**   There is evidence that the spatial distribution of population is an important determinant of redistributive pressures, particularly so in non-democratic countries (Campante and Do 2007). This can be related to the idea, as expressed in Ades and Glaeser (1995), that proximity to the capital city increases an individual's political influence. This is particularly the case with regard to "non-institutional" channels like demonstrations, insurgencies and revolutions, as opposed to democratic elections. As such, a more concentrated population is more capable of keeping a non-democratic government in check.[38] With that idea in the background, we study the correlation between our measures of concentration and a number of measures of the quality of governance, compiled by Kaufman, Kraay and Mastruzzi (2006). These results are featured in Table 5.

## [TABLE 5 HERE]

The first panel, for the full sample, displays a positive correlation between population concentration and governance. For conciseness, we present the results using the first principal component of five of the KKM variables (Control of Corruption, Voice & Accountability, Government Effectiveness, Rule of Law, and Regulation Quality).[39] Column (1) shows that higher concentration as measured by the G-CISC is associated with better governance. (We focus on $GCISC_1$.) The one measure of governance that displays a different pattern turns out to be Political Stability, as shown in Column (2), that shows no correlation. This is interesting in and

---

[37]One tentative way of probing deeper into this link with population size is to consider the effects of openness. Introducing openness into the regression reduces the coefficient and significance of population size, which may indicate that part of the negative relationship is indeed linked to the relative attraction of the capital city, which may be more pronounced in a more open, outward-oriented economy. The high correlation between openness and population makes it hard to disentangle their effects, however.

[38]See Traugott's (1995) account of the role of capital cities in revolutions, and how it relates to the absence of institutional outlets that could serve as alternatives to "insurrectionary politics".

[39]The measures are notoriously highly correlated, and the Kaiser-Meyer-Olkin measure of sampling adequacy is very high, always in excess of 0.85. The results for each individual measure are very similar, and are available upon request.

of itself. In fact, if we include the other governance variables as controls in a regression with stability as the dependent variable, as in Column (3), we see that population concentration has a *negative* and typically significant correlation with stability. This is consistent with the idea that, controlling for the quality of governance in non-democratic polities, the concentration of population around the capital city imposes checks on the incumbent government.[40] Columns (4)-(6) show that the same results hold when concentration is measured using the L-CISC.

An even more striking pattern emerges, however, when we split the sample between democracies and non-democracies: it is clear that this relationship is present only in non-democratic countries. In this sub-sample, a higher degree of concentration around the capital city is a strong predictor of higher governance quality with an increase of around 30% of standard deviation for an increase of one standard deviation in G-CISC. Essentially no effect is verified for more democratic countries, and the same pattern holds for the impact of concentration on political stability – both for G-CISC and L-CISC. In other words, non-democratic countries whose populations are more concentrated around the political center of power have better governance and, conditional on that level of governance, less political stability. This is precisely in line with the idea that the concentration of population represents a check on non-democratic governments, through the stability threat that it poses.

Table 5 already shows that the significance of the coefficients is generally improved with G-CISC, as opposed to L-CISC. This is not too surprising, in light of the more well-behaved distribution displayed by the former. In fact, we can further establish this comparison, while also considering how our measures of concentration fare when compared to the alternative measures we have been using as benchmarks. For that purpose, we run a "horse race" in which the measures are jointly included, as shown in Table 6 – for brevity, we only present the first principal component of the governance measures, and focus on the sample of non-democratic countries. It is clear that both G-CISC and L-CISC dominate the alternative measures, and that G-CISC seems to provide the clearest picture of the correlations linking the concentration of population around the capital city and governance.[41]

---

[40]These results, which are available upon request, are verified both when stability is measured by the Kaufman, Kraay and Mastruzzi (2006) index, and also when it is measured by the average length of tenure experienced by incumbent executives or parties in the previous twenty years. For details on this measure, see Campante, Chor, and Do (2009).

[41]One can once again note that the location Gini goes in the opposite direction of the centered measures.

[**TABLE 6 HERE**]

### 6.1.3   Where to Locate the Capital?

The idea that the capital city is a particularly important point from a political standpoint, and the correlation between the concentration of population around the capital and the extent of the checks on the government suggest that governments – and non-democratic ones in particular – would have an incentive to pick suitable locations for their capital. This draws attention to the endogeneity of the location of the capital city: not only is the concentration of population a variable that is determined in equilibrium, but the concentration patterns can also influence the choice of where to locate the capital. This is another idea that this application of our index enables us to address.

While a full treatment of causality is beyond the scope of this paper, we can nevertheless illustrate how our index can shed light on this topic. More generally, we can illustrate how our index helps approach the issue of the choice of the capital point of interest. Consider a country with a given spatial distribution of its population, and let us think of the problem faced by a ruler with respect to where to locate the country's capital.[42] There are centripetal forces that would lead the ruler to consider spots where the concentration would be very high – economies of agglomeration, broadly speaking. But there are other centrifugal forces, such as the afore-mentioned checks on his power, that would lead him to place the capital in a low-concentration spot. The question is, which of these forces will prevail under which circumstances?

Our index can provide an avenue for answering this question, which we illustrate using G-CISC. For every country, we compute the concentration of population around every single point in that country.[43] We then specify the point where this concentration reaches its maximum value. Interestingly, for three fourths of the countries (in the year 1990) this maximum-concentration location lies right within the capital city. This high rate is explained in part by the choice to put the capital in a central location, and in part by the fact that being the capital increases the location's attractiveness to migrants and to economic activity in general. More broadly, the maximum-concentration location is often at the largest city.[44]

---

[42]The history of changes in the location of capital cities, considered at some length in Campante and Do (2007), is proof that this problem is very often explicitly considered.

[43]More precisely, every single cell in the grid that covers the country.

[44]The exceptions are often illustrative. In China, it is close to Zhengzhou, the largest city in the province of Henan, which is the country's most populous; in India, similarly, it is in the state of Uttar Pradesh, which is also

We can then measure the gap between this site and the actual capital, as an indicator of how far a country's actual choice of capital is from the point that would maximize the "agglomeration economies". We regress this distance, normalized by the greatest distance to any point in the country, on a set of political variables using OLS and Tobit regressions. The results are presented in Table 7. When we limit ourselves to non-democratic countries, we see that a higher level of autocracy predicts a greater distance between the capital city and the concentration-maximizing location. Then when we limit ourselves to non-autocratic countries, then a higher level of democracy also predicts a greater distance. When combined together, both variables of autocracy and democracy predict a greater distance: this shows a type of U-shaped relationship, in which the centrifugal forces are strongest in both extremes of autocracy and democracy. This pattern is very robust to the inclusion of many dummy variables, including regional dummies and legal origin dummies. We can speculate that, on the autocratic side, more autocratic governments have greater incentive and/or ability to insulate themselves from popular pressure by locating their capital cities in low-concentration spots. On the democratic side, it is perhaps the case that additional democratic openness will lead to greater decentralization, and a lower level of attraction exerted by the capital. We are far from having a theory to fully account for that at this point, but the stylized fact is quite interesting nonetheless, and we also leave it to future research.

[**TABLE 7 HERE**]

## 6.2   US State-level Implementation

Building on this discussion on the location of the capital city, there is no better country in which to take our empirical implementation to the regional level than the United States, with its long tradition of dealing with the issue. Most famously, James Madison elaborated at length on the choice of the site of the capital city, during the 1789 Constitutional convention, arguing that one should "place the government in that spot which will be least removed from every part of the empire," and that "regard was also to be paid to the centre of population." He also pointed out that state capitals had sometimes been placed in "eccentric places," and that in

---

the most populous. In the US, it is Columbus, OH, right in the middle of the large population concentrations of the East Coast and the Midwest.

those cases "we have seen the people struggling to place it where it ought to be."[45]

The force identified by Madison have been very much at play in the case of US states, and our index enables us to get a snapshot of what the outcome has been. Regarding concentration normalized by the size of the territory ($GCISC_2$), shown in Table 8, Illinois is less concentrated around its capital Springfield than any country in the world. Even with country size, its level of concentration is still comparable with Canada, ranking 10 in the world. Not only for Illinois, but many US States where the capital is not the largest city have the level of concentration comparable to the least concentrated countries in the world. On the other hand, the amount of variation is comparable to that of the cross-country implementation, as the concentration in states such as Rhode Island and Hawaii is on a par with some of the most concentrated countries in the world. In general, the broad lessons from the cross-country level persist: $GCISC_1$ and $LCISC_1$ generate similar rankings, and the Gini captures a completely different concept.

[**TABLE 8 HERE**]

## 6.3   US Metropolitan Area Implementation

Finally, for the sake of completeness, we implement our indices in the context of US metropolitan areas. There is a large literature discussing and measuring urban sprawl in this context, to which we obviously will not do justice. In any event, the concept of sprawl is multidimensional, and many different measures have been used to capture these multiple elements (Galster et al. 2001, Glaeser and Kahn 2003). Our measure of spatial concentration captures what Galster et al. (2001) call "centrality", and following that paper we compute our indices taking city hall as our capital point of interest. The results for 24 major metropolitan areas are in Table 9.

[**TABLE 9 HERE**]

Some of the results are expected – the high levels of concentration in places such as Boston and New York – while others may be somewhat more surprising to the naked eye – such as the middling levels of concentration in Los Angeles, a city that is synonymous with urban sprawl, or perhaps the extremely low level of concentration in San Francisco.[46] From a broader

---

[45]These quotations were obtained from the website *The Founders' Constitution* (on Article 1, Section 8, Clause 17), available at http://press-pubs.uchicago.edu/founders/.

[46]Both results are very much consistent with the findings in Galster et al. (2001).

perspective, perhaps the most notable feature is how closely linked are the rankings for the three versions of CISC, and how numerically close are $GCISC_1$ and $GCISC_2$ compared to the country- and state-level contexts. The message is simple: when distances are relatively small, as they are bound to be for cities when compared to countries or states, differences in elasticity coefficients or normalization procedures are less important.[47]

# 7    Concluding Remarks

We have presented a general axiomatic approach to understanding the concept of spatial concentration around a point of interest. Based on a probabilistic interpretation of spatial distributions, we build an analogy with expected utility theory that yields an expected influence representation for concentration orders. We give such orders further content by imposing a couple of basic properties, monotonicity and rank invariance, and show that they pin down a specific class of measures, defined over very general spaces of economic interest. We then go on to illustrate the empirical implementation of the measure, and how this implementation highlights some of the advantages of our index over alternative approaches.

We emphasize that our approach is a very general one, and unapologetically so. Our idea was to build a common language to operationalize the concept of centered spatial concentration over a broad scope of applications, in geographical and also in more abstract spaces. We certainly hope that this language, and the measure that we obtain from it, can be widely applied. Empirically, the correlations that we are able to point out between our index and a number of variables of interest can be exploited further, with particular attention to issues of causality that are left outside the scope of this paper. An extension of our framework to an aggregate, non-centered measure of concentration is left to future promising research.

---

[47]It is important to keep in mind that we are talking about relative distances, and not their numerical value: scale does not matter for the CISC!

# References

[1] Ades, Alberto F. and Edward L. Glaeser (1995), "Trade and Circuses: Explaining Urban Giants," *Quarterly Journal of Economics* 110: 195-227.

[2] Anderson, James E. and Eric van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review* 93: 170-192.

[3] Aumann, Robert J. and Roberto Serrano (2008), "An Economic Index of Riskiness," *Journal of Political Economy*, 116: 810-836.

[4] Axler, Sheldon, Paul Bourdon and Wade Ramey (2001), *Harmonic Function Theory.* New York: Springer-Verlag.

[5] Busch, Marc L. and Eric Reinhardt (1999), "Industrial Location and Protection: The Political and Economic Geography of U.S. Nontariff Barriers," *American Journal of Political Science*, 43: 1028-1050.

[6] Campante, Filipe R., Davin Chor, and Quoc-Anh Do (2009), "Instability and the Incentives for Corruption," *Economics & Politics*, 21: 42-92.

[7] Campante, Filipe R. and Quoc-Anh Do (2007), "Inequality, Redistribution, and Population," Harvard Kennedy School Faculty Research WP Series RWP07-046..

[8] Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT), (2004), "Gridded Population of the World (GPW), Version 3", Columbia University. Available at http://beta.sedac.ciesin.columbia.edu/gpw.

[9] Clark, C. (1951), "Urban Population Densities," *Journal of the Royal Statistical Society* 114: 490-496.

[10] Collier, Paul and Anke Hoeffler (2004), "Greed and Grievance in Civil War," *Oxford Economic Papers*, 56: 563-595.

[11] Duclos, Jean-Yves, Joan Esteban, and Debraj Ray (2004), "Polarization: Concepts, Measurement, Estimation," *Econometrica* 72: 17371772.

[12] Ellison, Glenn and Edward L. Glaeser (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy* 105: 889-927.

[13] Esteban, Joan and Debraj Ray (1994), "On the Measurement of Polarization," *Econometrica* 62: 819-851.

[14] Fishburn, Peter C. (1975), "Unbounded Expected Utility, *The Annals of Statistics* 3:884-896.

[15] Fishburn, Peter C. (1976), "Unbounded Utility Functions in Expected Utility Theory, *Quarterly Journal of Economics* 90:163-168.

[16] Foster, James E. and Amartya K. Sen (1997), *On Economic Inequality (2nd ed.), Annexe*, Oxford: Oxford University Press.

[17] Foster, James E. and Anthony F. Shorrocks (1991), "Subgroup Consistent Poverty Indices," *Econometrica* 59: 687-709.

[18] Fryer, Roland G., Jr. and Richard Holden (2007), "Measuring the Compactness of Political Districting Plans," Harvard University mimeo.

[19] Fujita, Masahisa, Paul Krugman, and Anthony J. Venables (1999), *The Spatial Economy: Cities, Regions, and International Trade.* Cambridge, MA: MIT Press.

[20] Galster, George, Roy Hanston, Michael R. Ratcliffe, Harold Wolman, Stephen Coleman, and Jason Freihage (2001), "Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept," *Housing Policy Debate* 12: 681-717.

[21] Glaeser, Edward L. and Matthew Kahn (2004), "Sprawl and Urban Growth", in V. Henderson and J. Thisse (eds.), *The Handbook of Regional and Urban Economics*, Amsterdam: North Holland.

[22] Henderson, Vernon (2003), "The Urbanization Process and Economic Growth: The So-What Question," *Journal of Economic Growth* 8: 47-71.

[23] Hidalgo, C. A., B. Klinger, A.-L. Barabási, and R. Hausmann (2007), "The Product Space Conditions the Development of Nations," *Science* 317: 482-487.

[24] Jensen, J. Bradford and Lori Kletzer (2005), "Tradable Services: Understanding the Scope and Impact of Services Offshoring," *Brookings Trade Forum*, 6: 75-134.

[25] Karni, Edi and David Schmeidler (1991), "Utility Theory with Uncertainty, in W. Hildenbrand and H. Sonnenschein (eds.), *Handbook of Mathematical Economics*, Amsterdam: North Holland.

[26] Kreps, David M. (1988), "Notes on the Theory of Choice. Boulder, CO: Westview Press, Inc.

[27] Krugman, Paul R. (1991), *Geography and Trade*, Cambridge, MA: MIT Press.

[28] Ledyard, J. O. (1971), "A Pseudo-Metric Space of Probability Measures and the Existence of Measurable Utility, *The Annals of Mathematical Statistics* 42:794-798.

[29] McMillen, Daniel P. (2004), "Employment Densities, Spatial Autocorrelation, and Sub-centers in Large Metropolitan Areas," *Journal of Regional Science* 44: 225-243.

[30] Mills, Edwin S. (1970), "Urban Density Functions," *Urban Studies* 7: 5-20.

[31] Ransford, Thomas (1995), *Potential Theory in the Complex Plane.* Cambridge, UK and New York: Press Syndicate of the University of Cambridge.

[32] Rogers, Andrei and Stuart Sweeney (1998), "Measuring the Spatial Focus of Migration Patterns," *The Professional Geographer*, 50: 232-242.

[33] Traugott, Mark (1995), "Capital Cities and Revolution," *Social Science History* 19: 147-168.

# A  Appendix: Proofs and Discussion

## A.1  Theorem 2

Before proceeding with the proofs, we state the following basic result regarding Cauchy's functional equation:

**Lemma 1** *A function $f$ on $\mathbb{R}$, continuous at at least one point, and satisfying Cauchy's functional equation $f(x) + f(y) = f(x + y)$ must have the form $f(x) = ax$ for some constant $a$.*

**Proof of Theorem 2.**  We first prove that Axioms 1-3 impose the functional form stated in Theorem 2 in two steps, before verifying the converse. Following Theorem 1, we write $\mathbf{I_C}(p) = \int_X h_{\mathbf{C}}(\mathbf{x})dp(\mathbf{x})$, where $h$ is a continuous function on $X$.

Axiom 3 implies immediately that for any distributions $p, q \in \mathscr{P}_X$ if $p \sim q$ (i.e. $\mathbf{I_C}(p) = \mathbf{I_C}(q)$), then for all $\rho$ positive: $\mathcal{S}_{(\mathbf{C},\rho)}(p) \sim \mathcal{S}_{(\mathbf{C},\rho)}(q)$ and $\mathbf{I_C}(\mathcal{S}_{(\mathbf{C},\rho)}(p)) = \mathbf{I_C}(\mathcal{S}_{(\mathbf{C},\rho)}(q))$.

**Step 1: Determining $h_{\mathbf{C}}$ on a ray.**  Consider a ray $\theta$ originating from the capital $\mathbf{C}$. Within Step 1, for all points $\mathbf{z}$ on this ray $\theta$ we denote $h(z) = h_{\mathbf{C}}(\mathbf{z})$, where $z = |\mathbf{C} - \mathbf{z}|$ is the distance to the capital. Consider three points $\mathbf{x}$, $\mathbf{1}$ and $\mathbf{y}$ on the ray $\theta$, respectively at distances $|\mathbf{C} - \mathbf{x}| = x$, $|\mathbf{C} - \mathbf{1}| = 1$ and $|\mathbf{C} - \mathbf{y}| = y$ from $\mathbf{C}$. Assume that $x > 1 > y$.

**Monotonicity of $h$.**  Consider the simple distributions $p_{\mathbf{x}}$ and $p_{\mathbf{y}}$: by continuity of $h_{\mathbf{C}}$ we see that $p_{\mathbf{x}} \precsim_{\mathbf{C}} p_{\mathbf{y}} \Leftrightarrow h(x) \leq h(y)$. Similarly, after a squeeze of coefficient $\rho > 0$ towards $\mathbf{C}$ we obtain $p_{\mathcal{S}_{(\mathbf{C},\rho)}(\mathbf{x})} \precsim_{\mathbf{C}} p_{\mathcal{S}_{(\mathbf{C},\rho)}(\mathbf{y})} \Leftrightarrow h(\rho x) \leq h(\rho y)$. Besides, Axiom 3 implies that $p_{\mathcal{S}_{(\mathbf{C},\rho)}(\mathbf{x})} \precsim_{\mathbf{C}} p_{\mathcal{S}_{(\mathbf{C},\rho)}(\mathbf{y})} \Leftrightarrow p_{\mathbf{x}} \precsim_{\mathbf{C}} p_{\mathbf{y}}$; therefore $h(x) \leq h(y) \Leftrightarrow h(\rho x) \leq h(\rho y)$. As we could rewrite $h(x) = h(y)$ as $h(x) \geq h(y)$ and $h(x) \leq h(y)$, the inequalities becomes equalities at the same time.

Consider the simple distributions $p_{\rho \mathbf{x}}$ and $p_{\rho \mathbf{y}}$. $\rho x > \rho y$ implies $F_{p_{\rho \mathbf{x}}}^{\mathbf{C}} \succsim_{FOSD} F_{\rho \mathbf{y}}^{\mathbf{C}}$. When $\rho$ is small enough, Axiom 2 states that $p_{\rho \mathbf{x}} \precsim_{\mathbf{C}} p_{\rho \mathbf{y}}$, so $h(\rho x) < h(\rho y)$. It follows that $h(x) < h(y)$ when $x > y$, or $h(z)$ is a strictly decreasing function.

**Finding an invariant condition for $h$.**  From the monotonicity of $h$, there exists a real number $\lambda \in (0, 1)$ such that: $\lambda h(x) + (1 - \lambda)h(y) = h(1)$ (similarly, that $\lambda p_{\mathbf{x}} + (1 - \lambda)p_{\mathbf{y}} \sim p_{\mathbf{1}}$). Squeezing the simple distributions $p_{\mathbf{x}}$, $p_{\mathbf{1}}$ and $p_{\mathbf{y}}$ by a ratio $k$, Axiom 3 implies: $\lambda h(kx) + (1 - \lambda)h(ky) = h(k)$. Eliminating $\lambda$ in the last two equalities, we obtain:

$$\frac{h(kx) - h(ky)}{h(x) - h(y)} = \frac{h(ky) - h(k)}{h(y) - h(1)}.$$

This equation shows that the left hand side does not depend on $x$. Given the symmetric roles that $x$ and $y$ play, it is immediate that the left hand side term does not depend on $y$ either. It is thus a function of $k$ only:

$$\frac{h(kx) - h(ky)}{h(x) - h(y)} = g(k). \tag{1}$$

Notice that if $\frac{a}{b} = \frac{c}{d}$ then $\frac{a}{b} = \frac{c}{d} = \frac{a-c}{b-d}$. Now for $x, z > 1$, equation (1) implies $\frac{h(kz)-h(ky)}{h(z)-h(y)} = g(k)$, therefore $g(k) = \frac{(h(kx)-h(ky))-(h(kz)-h(ky))}{(h(x)-h(y))-(h(z)-h(y))} = \frac{h(kx)-h(kz)}{h(x)-h(z)}$. Analogously in the case $y, z < 1$, we deduce that equation (1) holds for any positive numbers $x, y$.

39

**Solving for $g(k)$.** Further investigation of equation (1) shows that $\forall k, l > 0$:

$$g(kl) = \frac{h(klx) - h(kly)}{h(x) - h(y)} = \frac{h(klx) - h(kly)}{h(lx) - h(ly)} \frac{h(lx) - h(ly)}{h(x) - h(y)} = g(k)g(l).$$

Defining $\tilde{g}(u) \overset{\text{def}}{\equiv} \log(g(e^u))$ for $u \in \mathbb{R}$, it follows that $\tilde{g}$ is continuous and satisfies the Cauchy functional equation, and according to Lemma 1 must be of the form $\tilde{g}(u) = \gamma u$, or equivalently, $g(k) = k^\gamma$.

**Differentiability of $h(x)$.** We focus on the case $\gamma > 0$, as the other case is analogous. We replace $(x, y)$ in equation (1) by $(k^n x, k^{n-1} x), n = 1, 2, \ldots$ and rearrange the terms to deduce by induction that:

$$h(k^n x) - h(k^{n-1} x) = k^\gamma (h(k^{n-1} x) - h(k^{n-2} x)) = \ldots = k^{\gamma(n-1)}(h(kx) - h(x)),$$

from which we obtain by adding up all the differences:

$$h(k^n x) - h(x) = \frac{k^{\gamma n} - 1}{k^\gamma - 1} \left( h(kx) - h(x) \right). \tag{2}$$

We reason by absurd: suppose that $h(x)$ is not differentiable at $x$, i.e. there exist two sequences $({}_1 x_i)$ and $({}_2 x_i)$ both converging to $x$, but producing two different limits of $\frac{h(x_i) - h(x)}{x_i - x}$. These two series must also produce two different limits of $\frac{h(x_i) - h(x)}{x_i^\gamma - x^\gamma} x^\gamma \overset{\text{def}}{\equiv} D\gamma(x_i, x) \leq 0$, denoted $D\gamma_1$ and $D\gamma_2$. Suppose $D\gamma_1 - D\gamma_2 = \epsilon > 0$. For convenience, we only consider the case where members of both sequences are greater than $x$ (the other case is identical, by a change of variable). Given any $\delta > 0$, there exist $k_1$ and $k_2$ in the interval $(1, 1 + \delta)$ such that $k_1 x \in ({}_1 x_i)$ and $k_2 x \in ({}_2 x_i)$. For any $\eta > 0$, we could find natural numbers $m$ and $n$ such that $k_2^{n\gamma} + \eta \geq k_1^{m\gamma} \geq k_2^{n\gamma}$ (by choosing a rational number $\frac{m}{n}$ bigger than, yet as close to $\frac{\log k_2}{\log k_1}$ as possible). From equation (2), we obtain: $h(k_1^m x) = D\gamma(k_1 x, x)(k_1^{\gamma m} - 1) + h(x)$ and $h(k_2^n x) = D\gamma(k_2 x, x)(k_2^{\gamma n} - 1) + h(x)$.

Here we could choose $\delta$ small enough so that $D\gamma(k_1 x, x)$ and $D\gamma(k_2 x, x)$ are very close to $D\gamma_1$ and $D\gamma_2$, such that $D\gamma(k_1 x, x) - D\gamma(k_2 x, x) > \frac{1}{2}\epsilon > 0$. It follows that:

$$
\begin{aligned}
0 > D\gamma(k_1 x, x)(k_1^{\gamma m} - 1) &> D\gamma(k_2 x, x)(k_2^{\gamma n} + \eta - 1) + \frac{1}{2}\epsilon(k_2^{\gamma n} + \eta - 1) \\
&= D\gamma(k_2 x, x)(k_2^{\gamma n} - 1) + \frac{1}{2}\epsilon(k_2^{\gamma n} + \eta - 1) + D\gamma(k_2 x, x)\eta \\
&> D\gamma(k_2 x, x)(k_2^{\gamma n} - 1),
\end{aligned}
$$

when we choose $\eta$ to be very small compared to $k_2^{\gamma n}$. The last inequality implies that $h(k_1^m x) > h(k_2^n x)$, in contradiction with $k_1^m > k_2^n$ and $h$ being a decreasing function. We have thus proved by absurd that $h$ is differentiable everywhere.

**Solving for $h(x)$.** From equation (1), by fixing $y = 1$ and bringing $x$ towards 1 we obtain: $h'(k) = k^{\gamma-1} h'(1)$. This differential equation yields the following solutions: $h(x) = \alpha x^\gamma + \beta$ or $h(x) = \alpha \log x + \beta$ in case $\gamma = 0$, with constants $\alpha, \beta$. As $h$ is strictly decreasing, we have $\alpha\gamma < 0$.

**Step 2: Determining $h_{\mathbf{C}}$.** From Step 1, we could write: $h_{\mathbf{C}}(\mathbf{x}) = \alpha(\theta)|\mathbf{x} - \mathbf{C}|^{\gamma(\theta)} + \beta(\theta)$ or $\alpha(\theta)\log|\mathbf{x} - \mathbf{C}| + \beta(\theta)$, where $\theta$ denotes the ray from $\mathbf{C}$ to $\mathbf{x}$, and $\alpha(\theta)\gamma(\theta) < 0$.

Take two different rays $\theta_1$ and $\theta_2$, corresponding to the parameters $\alpha_1, \beta_1, \gamma_1$ and $\alpha_2, \beta_2, \gamma_2$ (the proof is analogous for the case of the log function). Consider the points $\mathbf{x}$ and $\mathbf{y}$ respectively on $\theta_1$ and $\theta_2$, at distances $x$ and $y$ from $\mathbf{C}$, such that $h_{\mathbf{C}}(\mathbf{x}) = h_{\mathbf{C}}(\mathbf{y})$, or $\alpha_1 x^{\gamma_1} + \beta_1 = \alpha_2 y^{\gamma_2} + \beta_2$. Considering the squeeze $\mathcal{S}_{(\mathbf{C},\rho)}$, as $h_{\mathbf{C}}(\mathcal{S}_{(\mathbf{C},\rho)}(\mathbf{x})) = h_{\mathbf{C}}(\mathcal{S}_{(\mathbf{C},\rho)}(\mathbf{y})) \,\forall \rho > 0$, it follows that $\alpha_1 x^{\gamma_1} \rho^{\gamma_1} + \beta_1 = \alpha_2 y^{\gamma_2} \rho^{\gamma_2} + \beta_2$, or $(\alpha_1 x^{\gamma_1} \rho^{\gamma_1 - \gamma_2} - \alpha_2 y^{\gamma_2})\rho^{\gamma_2} + \beta_1 - \beta_2 = 0$ for all $\rho$. The last equation holds for all $\rho > 0$, therefore $\beta_1 = \beta_2$, $\gamma_1 = \gamma_2$ and $\alpha_1 x^{\gamma_1} = \alpha_2 y^{\gamma_2}$.

Consider the case $x > y$. It follows that $F_{p_{\mathbf{x}}}^{\mathbf{C}} \succ_{FOSD} F_{p_{\mathbf{y}}}^{\mathbf{C}}$, so Axiom 2 implies that $p_{\mathbf{x}} \prec p_{\mathbf{y}}$, from which we infer that $h_{\mathbf{C}}(\mathbf{x}) < h_{\mathbf{C}}(\mathbf{y})$. Now let $x \to y_+$, by continuity we obtain $h_{\mathbf{C}}(\mathbf{x}) \leq h_{\mathbf{C}}(\mathbf{y})$. Similarly, consider $x < y$ and let $x \to y_-$, we would deduce $h_{\mathbf{C}}(\mathbf{x}) \geq h_{\mathbf{C}}(\mathbf{y})$ when $x = y$. It follows that $h_{\mathbf{C}}(\mathbf{x}) = h_{\mathbf{C}}(\mathbf{y})$ whenever $x = y$, i.e. $\alpha_1 = \alpha_2$. Thus we could rewrite the influence function as:

$$h_{\mathbf{C}}(\mathbf{x}) = \alpha|\mathbf{x} - \mathbf{C}|^{\gamma} + \beta \ (\alpha\gamma < 0) \ \ or \ \ \alpha\log|\mathbf{x} - \mathbf{C}| + \beta \ (\alpha < 0).$$

**Step 3: Verifying Axioms 1-3.** The result of Theorem 1 assures that Axiom 1 is satisfied when $\gamma > 0$. When $\gamma < 0$ or $h(z) = \alpha\log z + \beta$, $h_{\mathbf{C}}$ is discontinuous at $\mathbf{C}$, therefore Axiom 1.1 is satisfied, but not Axiom 1.2. (This caveat of unbounded influence functions has been further discussed in the paper.)

Given two distributions $p$ and $q$ such that $F_p^{\mathbf{C}} \prec_{FOSD} F_q^{\mathbf{C}}$, because $h$ is strictly decreasing, we have:

$$\mathbf{I}_{\mathbf{C}}(p) = \int_X h(|\mathbf{x} - \mathbf{C}|)dp(\mathbf{x}) = \int_{\mathbb{R}^+} h(r)dF_p^{\mathbf{C}}(r) < \int_{\mathbb{R}^+} h(r)dF_q^{\mathbf{C}}(r) = \int_X h(|\mathbf{x} - \mathbf{C}|)dq(\mathbf{x}) = \mathbf{I}_{\mathbf{C}}(q),$$

so Axiom 2 is satisfied. Finally, for any pair of distributions $p, q \in \mathscr{P}_X$ such that $p \succsim_{\mathbf{C}} q$, or equivalently $\int_X h(|\mathbf{x} - \mathbf{C}|)dp(\mathbf{x}) \geq \int_X h(|\mathbf{x} - \mathbf{C}|)dq(\mathbf{x})$, Axiom 3 would be verified if the following inequality holds for all $\rho > 0$:

$$\int_X h(\rho|\mathbf{x} - \mathbf{C}|)dp(\mathbf{x}) \geq \int_X h(\rho|\mathbf{x} - \mathbf{C}|)dq(\mathbf{x}) \tag{3}$$

If $h(z) = az^{\alpha} + b$, (3) is equivalent to:

$$\int_X [\rho^{\alpha}(h(|\mathbf{x} - \mathbf{C}|) - b) + b]\, dp(\mathbf{x}) \geq \int_X [\rho^{\alpha}(h(|\mathbf{x} - \mathbf{C}|) - b) + b]\, dq(\mathbf{x})$$

$$\Leftrightarrow \rho^{\alpha} \int_X h(|\mathbf{x} - \mathbf{C}|)dp(\mathbf{x}) - b(\rho^{\alpha} - 1) \geq \rho^{\alpha} \int_X h(|\mathbf{x} - \mathbf{C}|)dq(\mathbf{x}) - b(\rho^{\alpha} - 1).$$

The last inequality holds because $\rho > 0$.

If $h(z) = a\log z + b$, analogously, (3) is equivalent to:

$$\int_X [h(|\mathbf{x} - \mathbf{C}|) + a\log\rho]\, dp(\mathbf{x}) \geq \int_X [h(|\mathbf{x} - \mathbf{C}|) + a\log\rho]\, dq(\mathbf{x})$$

$$\Leftrightarrow \int_X h(|\mathbf{x} - \mathbf{C}|)dp(\mathbf{x}) + a\log\rho \geq \int_X h(|\mathbf{x} - \mathbf{C}|)dq(\mathbf{x}) + a\log\rho,$$

which is automatically verified. This completes the proof of Theorem 2. ∎

## A.2    Proposition 1

**Proof of Proposition 1.**    By continuity of the influence function, Condition 1 implies: $h(0) + h(x) \geq h(\eta x) + h((1 - \eta)x)$, or:

$$h(x - \eta x) - h(x) \leq h(0) - h(\eta x), \tag{4}$$

for some very small $\eta$. From Theorem 2 we know that $h(z)$ is either convex or concave in the distance to the capital, because $h''(z)$ is either positive or negative over $\mathbb{R}^+$. The above inequality is inconsistent with $h(z)$ being concave, so $h(z)$ must be convex. As $h''(z) = \alpha\gamma(\gamma - 1)z^{\gamma-2}$, or $h''(z) = -\alpha z^{-2}$ in the log case, $h(z)$ is convex if and only if $R_h = 1 - \gamma \geq 0$. Conversely, if $R_h \geq 0$, $h(z)$ is convex, so inequality (4) is satisfied, thus so is Condition 1. So we have proved that Condition 1 is equivalent to $R_h \geq 0$. The equivalence between Condition 2 and $R_h \leq 0$ could be proved analogously. ∎

## A.3    Proposition 2

Proposition 2 requires some results in harmonic function theory, which we state here for convenience. They are discussed in much more detail, for instance, in Axler et al. (2001) and Ransford (1995).

**Definition 9 (Harmonic Function)** *A real function $f(x_1, x_2, \ldots, x_n)$ is said to be harmonic on an open domain $D$ of $\mathbb{R}^n$ if it satisfies the Laplace equation over that domain (with well-defined second order partial derivatives):*

$$\triangle f \stackrel{def}{=} \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \ldots + \frac{\partial^2 f}{\partial x_n^2} \equiv 0.$$

**Definition 10 (Super-/Sub-harmonic functions)** *Given $f(x_1, x_2, \ldots, x_n)$ a real-valued function on an open domain $D$ of $\mathbb{R}^n$ that has finite second order partial derivatives. It is said to be super-harmonic on $D$ if $\triangle f \leq 0$. It is said to be sub-harmonic if $-f$ is super-harmonic.*

In particular, a harmonic function transformed through a summation, scalar multiplication, translation, squeeze, rotation or partial/directional differentiation is still a harmonic function. The essential property of harmonic functions is its Mean Value Property.

**Definition 11** *Given a ball $\mathbf{B}(\mathbf{T}, \kappa)$ within the open domain $D \subset \mathbb{R}^n$, its sphere $\mathbf{S}(\mathbf{T}, \kappa) = \partial\mathbf{B}(\mathbf{T}, \kappa)$, and a harmonic function $f$ on $D$, $\sigma_{\mathbf{S}}$ as the uniform surface measure on the sphere, and $\nu_{\mathbf{B}}$ is the uniform (Lebesgue) measure on the ball. The mean of $f$ on the sphere and on the ball is defined as follows:*

$$\mathcal{M}_{\mathbf{S}(\mathbf{T},\kappa)}(f) \stackrel{def}{=} \frac{1}{\sigma_{\mathbf{S}}(\mathbf{S})} \int_{\mathbf{S}} f(x) d\sigma_{\mathbf{S}}; \quad \mathcal{M}_{\mathbf{B}(\mathbf{T},\kappa)}(f) \stackrel{def}{=} \frac{1}{\nu_{\mathbf{B}}(\mathbf{B})} \int_{\mathbf{B}} f(x) d\nu_{\mathbf{S}}$$

**Property 1 (Mean Value Property)** *(1) $f$ is harmonic on $D$ if and only if the mean of $f$ on the sphere and its value at the center of the sphere are equal:*

$$\mathcal{M}_{\mathbf{S}(\mathbf{T},\kappa)}(f) = f(\mathbf{T}).$$

*(2) $f$ is harmonic on $D$ if and only if the mean of $f$ on the ball and its value at the center of the ball are equal:*

$$\mathcal{M}_{\mathbf{B}(\mathbf{T},\kappa)}(f) = f(\mathbf{T}). \tag{5}$$

**Property 2 (Mean Value Inequality)** *(1) $f$ is super-harmonic (sub-harmonic) on $D$ if and only if for any sphere $\mathbf{S}(\mathbf{T}, \kappa)$ whose ball $\mathbf{B}(\mathbf{T}, \kappa)$ lies completely within the domain $D$, the following mean value inequality holds with a $\leq$ ($\geq$) sign:*

$$\mathcal{M}_{\mathbf{S}(\mathbf{T}, \kappa)}(f) \leq (\geq) f(\mathbf{T}),$$

*(2) $f$ is super-harmonic (sub-harmonic) on $D$ if and only if for any ball $\mathbf{B}(\mathbf{T}, \kappa) \subset D$, the following mean value inequality holds with a $\leq$ ($\geq$) sign:*

$$\mathcal{M}_{\mathbf{B}(\mathbf{T}, \kappa)}(f) \leq (\geq) f(\mathbf{T}), \tag{6}$$

The following Lemma is used for the proof of Proposition 2:

**Lemma 2** *Given a twice differentiable function $h(|\mathbf{x} - \mathbf{C}|)$ on $\mathbb{R}^n \backslash \{\mathbf{C}\}$, we have:*

$$\triangle h(|\mathbf{x} - \mathbf{C}|) = h''(|\mathbf{x} - \mathbf{C}|) + \frac{(n-1)}{|\mathbf{x} - \mathbf{C}|} h'(|\mathbf{x} - \mathbf{C}|).$$

**Proof of Lemma 2.** Straightforward algebra shows Lemma 2:

$$
\begin{aligned}
\triangle h(|\mathbf{x} - \mathbf{C}|) &= \sum_1^n \frac{\partial^2 h(|\mathbf{x} - \mathbf{C}|)}{\partial x_i^2} = \sum_1^n \frac{\partial \left( h' \cdot \frac{x_i}{|\mathbf{x} - \mathbf{C}|} \right)}{\partial x_i} \\
&= \sum_1^n \left( h'' \cdot \frac{x_i^2}{|\mathbf{x} - \mathbf{C}|^2} + h' \cdot \left( \frac{1}{|\mathbf{x} - \mathbf{C}|} - \frac{x_i^2}{|\mathbf{x} - \mathbf{C}|^3} \right) \right) \\
&= h''(|\mathbf{x} - \mathbf{C}|) + \frac{(n-1)}{|\mathbf{x} - \mathbf{C}|} h'(|\mathbf{x} - \mathbf{C}|).
\end{aligned}
$$

∎

**Proof of Proposition 2.**

We consider any ball $\mathbf{B}(\mathbf{T}, \kappa) \subset X \backslash \{\mathbf{C}\}$. Condition 4 implies the following inequality for any $\rho < 1$:

$$\mathcal{M}_{\mathbf{B}(\mathbf{T}, \kappa)}(h_{\mathbf{C}}) \leq \mathcal{M}_{\mathbf{B}(\mathbf{T}, \rho\kappa)}(h_{\mathbf{C}}) \tag{7}$$

Letting $\rho \to 0$, this inequality becomes the Mean Value Inequality in (6) for all points $\mathbf{T}$. Thus the influence function $h_{\mathbf{C}}(\mathbf{x}) \equiv h(|\mathbf{x} - \mathbf{C}|)$ must be super-harmonic on $X \backslash \{\mathbf{C}\}$, implying $\triangle h(|\mathbf{x} - \mathbf{C}|) \leq 0$. By Lemma 2, it follows that:

$$R_h = -\frac{|x - \mathbf{C}| h''(|x - \mathbf{C}|)}{h'(|x - \mathbf{C}|)} \leq n - 1.$$

Similarly, Condition 3 is equivalent to $\mathcal{M}_{\mathbf{B}(\mathbf{T}, \kappa)}(h_{\mathbf{C}}) \geq \mathcal{M}_{\mathbf{B}(\mathbf{T}, \rho\kappa)}(h_{\mathbf{C}})$, which would imply the opposite for $R_h$: $R_h \geq n - 1$.

Conversely, if $R_h \leq n - 1$, $h_{\mathbf{C}}$ must be super-harmonic on $X \backslash \{\mathbf{C}\}$, implying inequality (7). The proof for the sub-harmonic case is analogous. ∎

## A.4   The Problem of Unboundedness

Let us consider a simple example of a solution to the problem of unboundedness related to Theorem 2. We can restrict the set of probability distributions to that of simple probabilities, namely the set $\mathscr{P}_X^s$ of probabilities $p$ such that $p(A) = 1$ for a certain finite subset $A \subset X \backslash \{\mathbf{C}\}$. Furthermore, we can replace Axiom 1.2 with the Archimedean Axiom that for all $p, q, r \in$

$\mathscr{P}_X^s, p \succ q \succ r$, there exist real numbers $\alpha, \beta \in (0,1)$ such that $\alpha p + (1-\alpha)r \succ q \succ \beta p + (1-\beta)r$ (since Axiom 1.2 implies the Archimedean Axiom, the latter is less stringent). This modified Axiom 1 implies a result that is very much like Theorem 1, except for the boundedness and continuity of the function $h_{\mathbf{C}}$ (e.g. Fishburn 1975). On the other hand, influence functions from Theorem 2 with $R_h \geq 1$ now fully satisfy Axiom 1, and are thus admissible as solutions to Theorem 2.

A second approach is to restrict the set of locations $X$ to $X' = X \backslash \mathbf{B}(\mathbf{C}, \epsilon)^{int}$ for a very small $\epsilon$. In other words, we exclude a very small neighborhood of the capital from consideration. In practice, the interpretation is similar to the previous paragraph, and entails a negligible infinitesimal amputation of the space in consideration that would hardly be noticeable in any real world dataset. Under this small change, $X'$ remains compact, while the influence functions from Theorem 2 with $R_h \geq 1$ are now bounded in $X'$, thus satisfying the conditions in Theorem 1. Therefore, these functions are also admitted as solutions to Theorem 2.

Generally speaking, there is a range of mild restrictions on the set of probability distributions $\mathscr{P}$ that allows some influence functions with $R_h \geq 1$. Notably, we may consider the set $\mathscr{P} \subset \mathscr{P}_X$ of distributions $p$, such that $p$ is a convex combination of a Lebesgue probability $p_L$ on $X$ and a simple probability $p_S$, and that the density of $p_L$ is bounded above by a constant $\zeta$ within a small ball $\mathbf{B}(\mathbf{C}, \epsilon)$, and that $\mathscr{P}$ is closed under finite convex combinations. It is possible to prove that our system, with Continuity replaced by the Archimedean Axiom, satisfies Axioms 0-6 stated by Ledyard (1971), as long as the expected influence is finite. (This proof is available upon request.) This implies that influence functions with $R_h \geq 1$ are also solutions to Theorem 2, as long as $R_h < 2 + n$ where $n$ is the dimension of the space. (Intuitively speaking, this constraint keeps the expected influence finite for any probability in $\mathscr{P}$.)

## A.5 Generalization to Normed Vector Spaces

We have shown our results for a Euclidean space $\mathbb{R}^n$, or more precisely, a finite dimension vector space associated with the Euclidean metric. Our framework is easily generalizable to any normed vector space $(X, \|\|\|)$ where the vector space $X$ is defined with respect to the field of real numbers. Replacing the Euclidean distance by the distance derived from the norm $\|\|\|$, Axioms 1-3 are immediately adaptable to the new space. Theorem 1 holds thanks to Expected Utility Theory, and Theorem 2 holds when we require the continuity of the norm $\|\|\|$.

Here is a practical example where this generalization allows for more flexibility. Suppose that we are interested in the space of political, social and economic opinions, where each individual is represented by her opinions on several issues, numerated from 1 to $n$. The distance between two individuals could then be modeled more generally as a CES function of the form:

$$\|\mathbf{x} - \mathbf{y}\| = (a_1(x_1 - y_1)^\rho + \ldots + a_n(x_n - y_n)^\rho)^{\frac{1}{\rho}}.$$

This allows for any coefficient of substitutability between the coordinates, as well as a normalization coefficient $a_i$ for each coordinate. In such situation, our framework still permits the researcher to measure the spatial concentration of opinions around a focal point $\mathbf{C}$ (such as the political status quo).

## A.6 Generalization to Metric Spaces

We also conjecture, without proving rigorously, that Theorems 1 and 2 could be extended to a much more general context where one only needs to specify a well-behaved metric space, with no need for a vector space structure. A concrete example of a non-Euclidean metric is that of the time to travel between any two points on the map. It is a metric that satisfies the basic metric properties (identity of indiscernibles, symmetry and triangular inequality), while incorporating complications such as road conditions, natural and institutional barriers, etc.

We would assume the metric to be continuous in the topology used to define the space of distributions. (In the example of travel time, the travel-time metric should be continuous with respect to the Euclidean metric.) Axiom 1 remains well-defined in this generalized framework. In this metric space, we define the balls of center $\mathbf{C}$ as $\mathbf{B}(\mathbf{C}, \kappa) = \{\mathbf{x} : d(\mathbf{x}, \mathbf{C}) \leq \kappa\}$, with which Axiom 2 could immediately be adapted. Axiom 3 is trickier, as it requires an extension of the squeeze operation. We conjecture that with some conditions on the metric, the balls $\mathbf{B}(\mathbf{C}, \kappa)$ are homeomorphic, so we could define a family of homeomorphic transformations as "generalized squeezes towards $\mathbf{C}$", where each squeeze of ratio $\rho$ maps a point at distance $x$ from $\mathbf{C}$ to a point at distance $\rho x$ from $\mathbf{C}$.

We notice that Theorem 1 still holds in this generalized context, which we know directly from Expected Utility Theory. The proof of Theorem 2 would be somewhat more complicated, yet feasible. We can define the concept of a ray that passes through $\mathbf{C}$ and any point $\mathbf{x}$ as the set of images from the suitably defined squeezes of $\mathbf{x}$ towards $\mathbf{C}$ for all $\rho \in \mathbb{R}^+$, and build a similar proof to Theorem 2.

In sum, we could extend our analysis to an even more general setting, which could be tailored to specific aspects of the applications, while keeping the rigor and the properties guaranteed by the axiomatic approach.

# B Appendix: Data Description

**Population Concentration Index:** The measures $G - CISC$'s and $L - CISC$'s are calculated and normalized as explained in the text, using original gridded population maps from the database *Gridded Population of the World* (GPW), Version 3 from the Socio-Economic Data Center, Columbia University (2005), containing maps in 1990, 1995 and 2000 of a global grid of 2.5 arc-minute side cells (approximately 5km).

**Alternative Indices of Concentration:** The alternative indices of concentration are also produced from the same dataset. The location Gini (noted in the Tables as "Gini Pop") is calculated as the Gini coefficient of inequality of a special sample, in which each "individual" corresponds to a gridded cell on the map, and each individual's "income" corresponds to the size of the population living within that cell. "Cap Prim" (Capital city primacy) is calculated as the share of the capital city population over the total population. "Share Largest Point" and "Share Largest Urban Extent" are calculated as the ratio of respectively the largest settlement point and the largest urban extent over the total population. These population figures come directly from the SEDC.

**Gap to concentration maximizing location:** This variable is calculated for each country by measuring the distance between the actual site of the capital city, and the site of the capital that would maximize the G-CISC. The maximization is done with Matlab's large scale search method (with analytical gradient matrix), from a grid of 50 initial guesses evenly distributed on the country's map for large countries.

**Kaufmann, Kraay and Mastruzzi (KKM):** From KKM's (2006) indices, including Voice and Accountability, Control of Corruption, Rule of Law, Government Effectiveness, Political Stability, and Regulation Quality, themselves a composite of different agency ratings aggregated by an unobserved components methodology. On a scale of $-2.5$ to $2.5$. Data are available for 1996-2002 at two-year intervals, and thereafter for 2002-2005 on an annual basis. We use the data in 1996 for our measure of population concentration in 1990. KKM data available at: http://info.worldbank.org/governance/kkz2005/pdf/2005kkdata.xls

**Real GDP per capita:** From the World Bank World Development Indicators (WDI). Real PPP-adjusted GDP per capita (in constant 2000 international dollars).

**Population by year:** From the World Bank World Development Indicators (WDI).

**Democracy:** Polity IV democracy score, on a scale of 0 to 10.

**Autocracy:** Polity IV autocracy score, on a scale of 0 to 10.

**Polity:** Polity IV composite score as Democracy minus Autocracy, on a scale of -10 to 10.

The reference date for the annual observations in the Polity IV dataset is 31 December of each year. We match these to the data corresponding to 1 January of the following year for consistency with the DPI. Data available at: http://www.cidcm.umd.edu/inscr/polity/

**Ethno-Linguistic Fractionalization:** From Alesina et al. (2003).

**Legal Origin:** From La Porta et al. (1999). Dummy variables for British, French, Scandinavian, German, and socialist legal origin.

**Region dummies:** Following the World Bank's classifications, dummy variables for: East Asia and the Pacific; East Europe and Central Asia; Middle East and North America; South Asia; West Europe; North America; Sub-Saharan Africa; Latin America and the Caribbean.

## TABLE 1
### Cross Country Summary Statistics

| Variable | Observations | Mean | Standard Deviation | Min | Max |
|----------|--------------|------|--------------------|-----|-----|
| GCISC$_1$ 90 | 156 | 0.4639 | 0.0971 | 0.2455 | 0.7641 |
| GCISC$_1$ 95 | 156 | 0.4644 | 0.0971 | 0.2439 | 0.7641 |
| GCISC$_1$ 00 | 156 | 0.4648 | 0.0971 | 0.2418 | 0.7641 |
| GCISC$_2$ 90 | 156 | 0.2527 | 0.0737 | 0.1047 | 0.5820 |
| GCISC$_2$ 95 | 156 | 0.2534 | 0.0736 | 0.1004 | 0.5820 |
| GCISC$_2$ 00 | 156 | 0.2540 | 0.0737 | 0.0973 | 0.5820 |
| LCISC$_1$ 90 | 156 | 0.9678 | 0.0295 | 0.8347 | 0.9989 |
| LCISC$_1$ 95 | 156 | 0.9679 | 0.0295 | 0.8326 | 0.9989 |
| LCISC$_1$ 00 | 156 | 0.9679 | 0.0295 | 0.8301 | 0.9989 |
| Gini Pop 90 | 156 | 0.6496 | 0.1588 | 0.1388 | 0.9869 |
| Gini Pop 95 | 156 | 0.6515 | 0.1580 | 0.1244 | 0.9872 |
| Gini Pop 00 | 156 | 0.6538 | 0.1569 | 0.1097 | 0.9877 |
| Capital Primacy 90 | 110 | 0.1226 | 0.1268 | 0.0016 | 1.0337 |
| Capital Primacy 95 | 156 | 0.1174 | 0.1182 | 0.0011 | 1.1016 |
| Capital Primacy 00 | 156 | 0.1210 | 0.1153 | 0.0011 | 1.0241 |
| GCISC$_1$ Growth 90-95 | 156 | 0.0005 | 0.0026 | -0.0129 | 0.0109 |
| GCISC$_1$ Growth 95-00 | 156 | 0.0004 | 0.0023 | -0.0069 | 0.0116 |

**TABLE 2**
**Cross Country Correlation**

| | $GCISC_1$ 90 | $GCISC_1$ 95 | $GCISC_1$ 00 | $GCISC_2$ 90 | $GCISC_2$ 95 | $GCISC_2$ 00 | $LCISC_1$ 90 | $LCISC_1$ 95 | $LCISC_1$ 00 | Gini Pop 90 | Gini Pop 95 | Gini Pop 00 | Cap Prim 90 | Cap Prim 95 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **$GCISC_1$ 90** | 1 | | | | | | | | | | | | | |
| **$GCISC_1$ 95** | 0.9997 | 1 | | | | | | | | | | | | |
| **$GCISC_1$ 00** | 0.999 | 0.9997 | 1 | | | | | | | | | | | |
| **$GCISC_2$ 90** | 0.6326 | 0.6314 | 0.6298 | 1 | | | | | | | | | | |
| **$GCISC_2$ 95** | 0.6352 | 0.6351 | 0.6346 | 0.9988 | 1 | | | | | | | | | |
| **$GCISC_2$ 00** | 0.636 | 0.6369 | 0.6375 | 0.9953 | 0.9988 | 1 | | | | | | | | |
| **$LCISC_1$ 90** | 0.8453 | 0.845 | 0.8442 | 0.3548 | 0.3573 | 0.3584 | 1 | | | | | | | |
| **$LCISC_1$ 95** | 0.846 | 0.846 | 0.8454 | 0.3547 | 0.3576 | 0.3591 | 0.9998 | 1 | | | | | | |
| **$LCISC_1$ 00** | 0.8463 | 0.8465 | 0.8461 | 0.3542 | 0.3575 | 0.3595 | 0.9993 | 0.9998 | 1 | | | | | |
| **Gini Pop 90** | -0.2678 | -0.2718 | -0.2754 | 0.3652 | 0.3555 | 0.3455 | -0.4054 | -0.408 | -0.4099 | 1 | | | | |
| **Gini Pop 95** | -0.2699 | -0.2732 | -0.276 | 0.3662 | 0.3581 | 0.3499 | -0.4072 | -0.4096 | -0.4114 | 0.9987 | 1 | | | |
| **Gini Pop 00** | -0.2708 | -0.2733 | -0.2753 | 0.3663 | 0.3601 | 0.3537 | -0.4074 | -0.4096 | -0.4112 | 0.9942 | 0.9984 | 1 | | |
| **Cap Prim 90** | 0.4807 | 0.4792 | 0.4775 | 0.3814 | 0.3787 | 0.3749 | 0.3081 | 0.309 | 0.3096 | -0.073 | -0.0748 | -0.0773 | 1 | |
| **Cap Prim 95** | 0.4751 | 0.4739 | 0.4724 | 0.3733 | 0.371 | 0.3677 | 0.2995 | 0.3004 | 0.301 | -0.0807 | -0.0825 | -0.0848 | 0.9979 | 1 |
| **Cap Prim 00** | 0.4746 | 0.4736 | 0.4724 | 0.3855 | 0.3837 | 0.3809 | 0.3033 | 0.3044 | 0.3052 | -0.0754 | -0.0765 | -0.0782 | 0.9961 | 0.9979 |

# Table 3: Ranking by GCISC₁ 90

| Code | Country | GCISC$_1$ 90 | Rank GCISC$_1$ 90 | GCISC$_2$ 90 | Rank GCISC$_2$ 90 | LCISC$_1$ 90 | Rank LCISC$_1$ 90 | Gini Pop 90 | Rank Gini Pop 90 | Capital Primacy 95 | Rank Capital Primacy 95 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| USA | United States | 0.2455 | 1 | 0.2455 | 74 | 0.8347 | 1 | 0.9139 | 149 | 0.0022 | 4 |
| BRA | Brazil | 0.2467 | 2 | 0.1471 | 12 | 0.8809 | 6 | 0.8518 | 140 | 0.0116 | 12 |
| CHN | China | 0.2511 | 3 | 0.1688 | 21 | 0.8760 | 4 | 0.7507 | 113 | 0.0085 | 9 |
| ZAF(b) | South Africa (Cape Town) | 0.2631 | 4 | 0.1047 | 1 | 0.8774 | 5 | 0.9230 | 150 | 0.0522 | 45 |
| RUS | Russian | 0.2691 | 5 | 0.2501 | 77 | 0.8388 | 2 | 0.9298 | 153 | 0.0636 | 52 |
| IND | India | 0.2704 | 6 | 0.1708 | 22 | 0.8895 | 7 | 0.5405 | 39 | 0.0097 | 10 |
| MOZ | Mozambique | 0.2900 | 7 | 0.1449 | 11 | 0.8918 | 8 | 0.6613 | 88 | 0.0524 | 47 |
| KAZ | Kazakhstan | 0.2983 | 8 | 0.1487 | 14 | 0.9165 | 11 | 0.7502 | 112 | 0.0190 | 18 |
| ZAR | Congo Kinshasa (DR) | 0.2985 | 9 | 0.1563 | 15 | 0.8928 | 9 | 0.6063 | 58 | 0.0822 | 73 |
| CAN | Canada | 0.3014 | 10 | 0.2435 | 72 | 0.8726 | 3 | 0.9869 | 156 | 0.0264 | 25 |
| | | | | | | | | | | | |
| PRI | Puerto Rico | 0.6215 | 147 | 0.3541 | 146 | 0.9948 | 150 | 0.4932 | 21 | 0.1063 | 91 |
| SLV | El Salvador | 0.6283 | 148 | 0.3446 | 142 | 0.9946 | 148 | 0.5312 | 37 | 0.0794 | 69 |
| CRI | Costa Rica | 0.6315 | 149 | 0.3919 | 152 | 0.9939 | 145 | 0.6543 | 86 | 0.0752 | 62 |
| ARM | Armenia | 0.6446 | 150 | 0.4040 | 154 | 0.9948 | 151 | 0.5645 | 50 | 0.3317 | 153 |
| TTO | Trinidad and Tobago | 0.6478 | 151 | 0.3458 | 144 | 0.9962 | 154 | 0.6136 | 62 | 0.0367 | 31 |
| LBN | Lebanon | 0.6484 | 152 | 0.3283 | 137 | 0.9954 | 152 | 0.5957 | 56 | 0.3259 | 151 |
| JOR | Jordan | 0.6520 | 153 | 0.4501 | 155 | 0.9947 | 149 | 0.8840 | 147 | 0.2411 | 141 |
| KWT | Kuwait | 0.6653 | 154 | 0.3841 | 149 | 0.9963 | 155 | 0.7319 | 104 | 0.0170 | 15 |
| MUS | Mauritius | 0.7038 | 155 | 0.5820 | 156 | 0.9961 | 153 | 0.6268 | 70 | 0.1082 | 93 |
| SGP | Singapore | 0.7641 | 156 | 0.3529 | 145 | 0.9989 | 156 | 0.5162 | 28 | 1.1016 | 156 |

# Table 4: Predictors of Population Concentration

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Dependent variable → | $GCISC_1$ | | $GCISC_2$ | | $LCISC_1$ | | Gini Pop | |
| Log Population | -0.00831** | -0.00951** | -0.008 | -0.00989* | -0.00456*** | -0.00620*** | -0.0186* | -0.002 |
| | [0.00393] | [0.00385] | [0.00610] | [0.00582] | [0.00159] | [0.00156] | [0.0102] | [0.0127] |
| Log Land Area | -0.0476*** | -0.0465*** | -0.0145*** | -0.0147** | -0.0118*** | -0.00884*** | 0.0587*** | 0.0446*** |
| | [0.00262] | [0.00264] | [0.00531] | [0.00573] | [0.00153] | [0.00104] | [0.00844] | [0.00940] |
| Log GDP per capita | 0.003 | 0.005 | 0.0121** | 0.010 | -0.002 | -0.00526* | 0.0566*** | 0.0750*** |
| | [0.00354] | [0.00871] | [0.00591] | [0.0131] | [0.00186] | [0.00309] | [0.0104] | [0.0240] |
| Polity Score | | -0.001 | | 0.001 | | 0.000 | | 0.004 |
| | | [0.000732] | | [0.00144] | | [0.000261] | | [0.00234] |
| Ethno-Linguistic Frac | | -0.024 | | -0.038 | | -0.004 | | 0.129* |
| | | [0.0201] | | [0.0325] | | [0.00630] | | [0.0654] |
| Region FE | | YES | | YES | | YES | | YES |
| Legal Origin FE | | YES | | YES | | YES | | YES |
| Observations | 113 | 108 | 113 | 108 | 113 | 108 | 113 | 108 |
| R-squared | 0.817 | 0.868 | 0.222 | 0.462 | 0.657 | 0.783 | 0.421 | 0.578 |

*** p<0.01, ** p<0.05, * p<0.1

Intercepts omitted. Robust standard errors in brackets. All independent variables are taken with a 5-year lag.
G-CISC and L-CISC are both for 1990.

See Appendix for description of variables and sources.

# Table 5: Governance and Population Concentration

| Dependent variable → | (1) Governance (First Princ. Comp.) | (2) Political Stability | (3) Political Stability | (4) Governance (First Princ. Comp.) | (5) Political Stability | (6) Political Stability |
|---|---|---|---|---|---|---|
| **A. Full Sample** | | | | | | |
| $GCISC_1$ | 2.698** [1.081] | -0.378 [0.825] | -1.426* [0.729] | | | |
| LCISC | | | | 6.554** [3.189] | -1.305 [2.874] | -3.62 [2.463] |
| Control for Governance | | | Yes | | | Yes |
| Obs | 128 | 134 | 128 | 128 | 134 | 128 |
| R2 | 0.88 | 0.541 | 0.671 | 0.876 | 0.541 | 0.666 |
| **B. More Democratic Countries** | | | | | | |
| $GCISC_1$ | -0.016 [1.327] | 0.148 [1.134] | -0.194 [0.953] | | | |
| LCISC | | | | -1.488 [3.698] | -0.441 [3.730] | -1.123 [3.006] |
| Control for Governance | | | Yes | | | Yes |
| Obs | 78 | 80 | 78 | 78 | 80 | 78 |
| R2 | 0.888 | 0.675 | 0.797 | 0.888 | 0.675 | 0.797 |
| **C. Less Democratic Countries** | | | | | | |
| $GCISC_1$ | 6.227*** [1.609] | -1.083 [1.461] | -4.497** [1.732] | | | |
| LCISC | | | | 15.84** [7.224] | 3.274 [6.168] | -0.546 [6.661] |
| Control for Governance | | | Yes | | | Yes |
| Obs | 50 | 54 | 50 | 50 | 54 | 50 |
| R2 | 0.769 | 0.387 | 0.619 | 0.731 | 0.385 | 0.551 |

Intercept omitted. Robust standard errors in brackets. All regressions include Log GDP per capita, Log Population, Polity, and Ethno-Linguistic Fractionalization, plus Region and Legal Origin FEs. Panel B consists of countries with polity score larger than 5, Panel C consists of countries with polity score less than or equal to 5. GCISC and LCISC are both for 1990. Independent variables are taken with lag. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

**Table 6: Governance and Population Concentration in Less Democratic Countries**

| Dependent variable:<br>Governance (First Princ. Comp.) | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| $GCISC_1$ 90 | 5.967*** | 6.420*** | 6.355*** | | |
| | [1.940] | [1.821] | [2.104] | | |
| $LCISC_1$ 90 | 1.398 | | | 16.05** | 15.25** |
| | [6.909] | | | [7.015] | [7.339] |
| Gini Pop 90 | | -0.425 | | 0.426 | |
| | | [1.004] | | [0.899] | |
| Cap Prim 95 | | | -0.148 | | 0.865 |
| | | | [0.912] | | [0.889] |
| Observations | 50 | 50 | 50 | 50 | 50 |
| R-squared | 0.769 | 0.77 | 0.769 | 0.733 | 0.738 |

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Intercepts omitted. Robust standard errors in brackets. All regressions include Log GDP per capita, Log Population, Polity, and Ethno-Linguistic Fractionalization, plus Region and Legal Origin FEs. All independent variables are taken with a 5-year lag.

See Appendix for description of variables and sources.

## Table 7: Capital City and gap to the G-CISC-maximizing location

| Dependent variable: Gap Distance (1990) | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | A. Democratic Countries | | B. Autocratic Countries | | C. Full Sample | | |
| | Tobit | OLS | Tobit | OLS | Tobit | OLS | OLS |
| Autocracy | 0.137*** | 0.025** | | | 0.165*** | 0.038*** | 0.040*** |
| | [0.048] | [0.010] | | | [0.040] | [0.012] | [0.012] |
| Democracy | | | 0.159* | 0.022** | 0.143*** | 0.023*** | 0.032*** |
| | | | [0.094] | [0.0088] | [0.038] | [0.011] | [0.012] |
| Log(Population) | 0.044 | 0.0297** | 0.100** | 0.012 | 0.079*** | 0.032*** | 0.033* |
| | [0.045] | [0.013] | [0.045] | [0.024] | [0.029] | [0.012] | [0.019] |
| Log(GDP per Capita) | -0.092 | 0.041 | -0.098 | -0.039 | -0.082 | -0.002 | 0.001 |
| | [0.092] | [0.050] | [0.076] | [0.055] | [0.056] | [0.032] | [0.038] |
| Regional Fixed Effect | | YES | | YES | | YES | YES |
| Legal Origin Fixed Effect | | YES | | YES | | YES | YES |
| Other controls | | | | | | | YES |
| Observations | 58 | 58 | 54 | 54 | 113 | 113 | 100 |
| R-squared | . | 0.26 | . | 0.25 | . | 0.22 | 0.26 |

*** p<0.01, ** p<0.05, * p<0.1. Intercept omitted. Robust standard errors in brackets. Panel A consists of countries with autocracy score less than 2, Panel B consists of countries with democracy score les than 2. Dependent variable is calculated as the distance between the actual and the hypothetical capital city that maximizes G-CISC. Other control variables include Log Land Area, Landlock, Island, Coastshare (coastal line as proportion of total boundary). Independent variables are taken with lag 5. Tobit regressions' standard errors are bootstrapped with 500-replications.

## Table 8: Population Concentration in US States

| State | Code | GCISC_1 | Rank GCISC_1 | GCISC_2 | Rank GCISC_2 | LCISC | Rank LCISC | Gini Pop | Rank Gini Pop |
|---|---|---|---|---|---|---|---|---|---|
| **Top 10** | | | | | | | | | |
| Illinois | IL | 0.3029 | 6 | 0.0793 | 1 | 0.8916 | 8 | 0.3066 | 43 |
| South Dakota | SD | 0.3042 | 7 | 0.0943 | 2 | 0.8901 | 7 | 0.2945 | 36 |
| Florida | FL | 0.2309 | 2 | 0.0998 | 3 | 0.8068 | 2 | 0.3867 | 50 |
| Nevada | NV | 0.2801 | 4 | 0.1445 | 4 | 0.8182 | 4 | 0.3311 | 47 |
| Missouri | MO | 0.3447 | 10 | 0.1452 | 5 | 0.9215 | 15 | 0.2880 | 29 |
| Alaska | AK | 0.1465 | 1 | 0.1465 | 6 | 0.5801 | 1 | 0.2683 | 20 |
| Delaware | DE | 0.5101 | 43 | 0.1530 | 7 | 0.9759 | 46 | 0.2883 | 30 |
| New York | NY | 0.3171 | 8 | 0.1530 | 8 | 0.9007 | 9 | 0.3025 | 40 |
| Alabama | AL | 0.3625 | 12 | 0.1542 | 9 | 0.9230 | 16 | 0.3088 | 44 |
| California | CA | 0.2590 | 3 | 0.1617 | 10 | 0.8076 | 3 | 0.3089 | 45 |
| | | | | | | | | | |
| **Bottom 10** | | | | | | | | | |
| Oregon | OR | 0.4327 | 33 | 0.3081 | 41 | 0.9467 | 35 | 0.2805 | 27 |
| Maryland | MD | 0.5010 | 41 | 0.3114 | 42 | 0.9749 | 45 | 0.2922 | 33 |
| Nebraska | NE | 0.4454 | 36 | 0.3377 | 43 | 0.9406 | 29 | 0.2041 | 1 |
| Massachusetts | MA | 0.5548 | 48 | 0.3526 | 44 | 0.9779 | 49 | 0.2596 | 15 |
| Minnesota | MN | 0.4978 | 40 | 0.3830 | 45 | 0.9569 | 37 | 0.2519 | 10 |
| Arizona | AZ | 0.5088 | 42 | 0.3830 | 46 | 0.9580 | 39 | 0.3477 | 48 |
| Colorado | CO | 0.5261 | 45 | 0.3997 | 47 | 0.9642 | 40 | 0.2863 | 28 |
| Utah | UT | 0.5411 | 47 | 0.4238 | 48 | 0.9674 | 42 | 0.2656 | 18 |
| Rhode Island | RI | 0.6907 | 50 | 0.4465 | 49 | 0.9926 | 50 | 0.2674 | 19 |
| Hawaii | HI | 0.5855 | 49 | 0.4606 | 50 | 0.9663 | 41 | 0.3049 | 42 |

## Table 9: Population Concentration in US Metropolitan Statistical Areas

| city_id | City | GCISC_1 | Rank GCISC_1 | GCISC_2 | Rank GCISC_2 | LCISC | Rank LCISC | Gini Pop | Rank Gini Pop |
|---|---|---|---|---|---|---|---|---|---|
| 20 | San Francisco | 0.2169 | 1 | 0.1847 | 1 | 0.6626 | 1 | 0.3089 | 15 |
| 19 | San Diego | 0.2212 | 2 | 0.2212 | 2 | 0.6876 | 2 | 0.3089 | 16 |
| 10 | Houston | 0.2571 | 3 | 0.2485 | 3 | 0.7351 | 3 | 0.3179 | 19 |
| 8 | Dallas | 0.2607 | 4 | 0.2336 | 4 | 0.7422 | 4 | 0.3179 | 20 |
| 17 | Pittsburgh | 0.2724 | 5 | 0.1585 | 5 | 0.7792 | 7 | 0.2713 | 5 |
| 12 | Miami | 0.2764 | 6 | 0.2479 | 6 | 0.7637 | 6 | 0.3867 | 23 |
| 18 | Riverside | 0.2785 | 7 | 0.2665 | 7 | 0.7630 | 5 | 0.3089 | 17 |
| 5 | Cincinnati | 0.2795 | 8 | 0.1680 | 8 | 0.8103 | 9 | 0.2739 | 7 |
| 23 | Tampa | 0.2901 | 9 | 0.2112 | 9 | 0.8246 | 10 | 0.3867 | 24 |
| 11 | Los Angeles | 0.3282 | 10 | 0.3125 | 10 | 0.7909 | 8 | 0.3089 | 18 |
| 7 | Cleveland | 0.3319 | 11 | 0.2000 | 11 | 0.8587 | 13 | 0.2739 | 8 |
| 22 | St. Louis | 0.3368 | 12 | 0.2458 | 12 | 0.8332 | 11 | 0.2880 | 10 |
| 15 | Philadelphia | 0.3541 | 13 | 0.2635 | 13 | 0.8433 | 12 | 0.2713 | 6 |
| 6 | Detroit | 0.4016 | 14 | 0.3624 | 14 | 0.8944 | 15 | 0.2605 | 4 |
| 1 | Atlanta | 0.4075 | 15 | 0.3070 | 15 | 0.9035 | 16 | 0.3547 | 22 |
| 21 | Seattle | 0.4286 | 16 | 0.3338 | 16 | 0.9078 | 18 | 0.2527 | 2 |
| 4 | Chicago | 0.4420 | 17 | 0.3750 | 17 | 0.9057 | 17 | 0.3066 | 14 |
| 14 | New York | 0.4491 | 18 | 0.3732 | 18 | 0.8887 | 14 | 0.3025 | 12 |
| 16 | Phoenix | 0.4621 | 19 | 0.3789 | 19 | 0.9238 | 19 | 0.3477 | 21 |
| 13 | Minneapolis | 0.4664 | 20 | 0.3940 | 20 | 0.9240 | 20 | 0.2519 | 1 |
| 9 | Denver | 0.4857 | 21 | 0.3992 | 21 | 0.9347 | 21 | 0.2863 | 9 |
| 2 | Baltimore | 0.5130 | 22 | 0.3734 | 22 | 0.9610 | 23 | 0.2922 | 11 |
| 3 | Boston | 0.5144 | 23 | 0.3509 | 23 | 0.9589 | 22 | 0.2596 | 3 |
| 24 | Washington DC | 0.7397 | 24 | 0.3609 | 24 | 0.9937 | 24 | 0.3041 | 13 |