Supporting Information for "A Global Typology of Urban Energy Use and Potentials for an Urbanization Mitigation Wedge"

Felix Creutzig, Giovanni Baiocchi, Robert Bierkandt,
Peter-Paul Pichler, Karen C. Seto

Contents

1 Data

Data was used from three different sources: the World Bank [1, 2], which includes 45 cities, with data referring approximately to the year 2005, the Global Energy Assessment, GEA [3], which includes 225 cities, with data approximately referring to the year 2000, and the International Association of Public Transport, UITP [4], which includes 100 cities, and data from 1995. Cities are approximately representative of cities world wide in terms of population size (Fig. S1), demonstrating a log-log-linearly with population size. The slope between log city rank and log city size is -1.07, and the 95% confidence interval is between -1.21 and -0.93, including the value -1. Only those cities from the World Bank database were used where emissions due to aviation and marine could be separated; this yielded 26 cities. From the UITP database, we only used cities, which had a complete dataset on energy use, which was 87. All three databases include population data. UITP includes over 200 indicators on traffic, but used were only the metropolitan GDP per capita, population, population density and transport energy use to avoid overfitting in the regression analysis. The GEA database provides GDP per capita (PPP) data from Eurostat (2008) and PriceWaterhouseCoopers (2007). The GDP per capita data (PPP) for World Bank cities are from GEA, PriceWaterhouseCoopers (2009) [5] and Urban Audit [6]. We added to all three datasets data on heating and cooling degree days [7], a binary proxy variable for coastal location (researched), diesel and gasoline prices [8], household sizes [6], and the "Centers of Commerce Index" [9]. Urban population densities for World Bank and GEA observations were obtained from individual municipal sources. The latter data were not all consistent in terms of the definition of municipal boundaries, but were included nonetheless as a crude indication of population density. For complete statistics and description of the data used see Dataset S1.

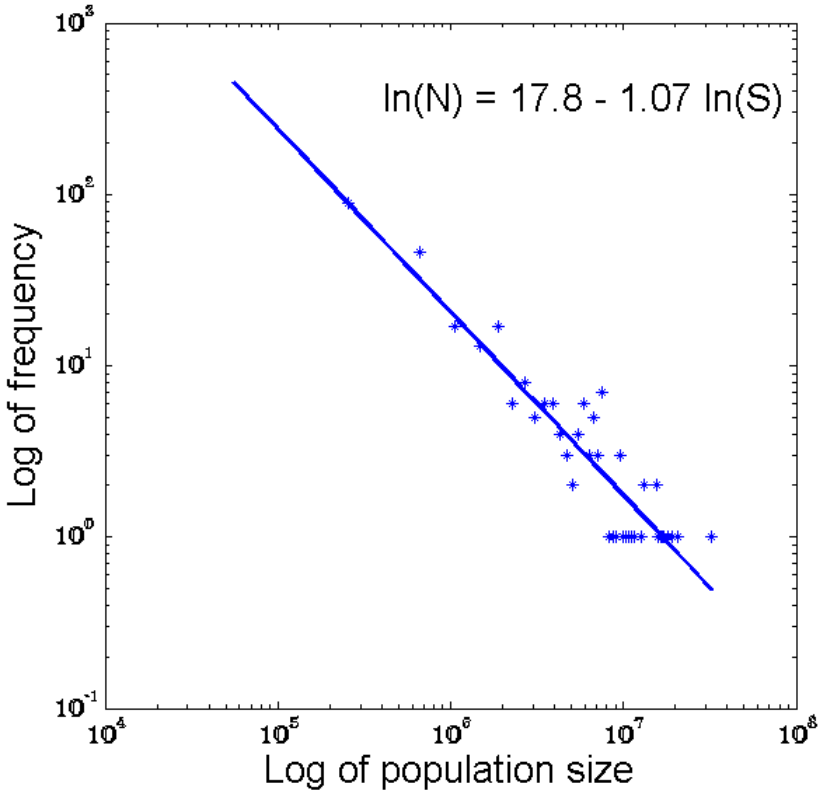


Figure S1: Rank-size statistics of cities analyzed. The sample includes cities of varying sizes, including global cities, and represent 21% of the global urban population.

2 Correlation analysis

Tab. 1 displays the Pearson's correlation coefficient between two variables X and Y defined as

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}.$$

Where $\mathrm{cov}(X,Y)$ represents the covariance, $\mathrm{var}(X)$ the variance of X and $\mathrm{var}(Y)$ the variance of Y. The Pearson correlation coefficient measures the linear relationship between two datasets, assuming that each dataset be normally distributed. -1 or $+1$ imply an exact linear relationship. Positive correlations indicate that as x increases, so does y, while negative correlations indicate that as x increases, y decreases. The p-value can be interpreted as the probability of an uncorrelated system producing datasets that have a Pearson correlation at least as large in magnitude as the one computed from these datasets. Partial correlation denotes the degree of association between two random variables, with the effect of a set of controlling random variables removed, in this case the respective other independent variables. To test for significance across all data sets and to synthesize heterogeneous research, we performed a meta-analysis, which essentially calculates the weighted average of the effect sizes of a group of studies, relying on the random effect DerSimonian-Laird (DSL) approach (see meta-analysis).

Table S1 presents the correlations statistics underlying Figure 1 in the main body text.

Table S2 provides cross correlation coefficients between transport costs and density metrics, indicating that transport costs could be a driving factor of urban form, and that thus gasoline prices not only have a direct influence but also an indirect influence on urban energy use.

Table S1: Correlation analysis, presenting Pearson's correlation coefficient. For the meta-analysis of Pearson correlation coefficient and also partial correlation coefficient the DerSimonian-Laird approach was used. WB: World Bank data; GEA: Global Energy Assessment; UITP. The binary variable for coastal location did not show any significant correlation with energy consumption. Significance levels: ***p<0.01; **p<0.05; *p<0.1.

	Dependent Variable	Number of cities	GDP/cap	Population density	HDD	CDD	Gasoline Price	Population	Household size	Urbanization level	Commerce center index
WB	GHGe/cap	26	0.65***	-0.60***	0.64***		-0.70***	-0.51***	-0.46**		
GEA	Final energy consumption/cap	225	0.26***	-0.12*	0.22***	-0.11*		-0.17***	-0.25***	0.26***	0.27*
UITP	Urban transport energy/cap	87	0.40***	-0.52***			-0.45***			0.41***	
Meta: Pearson		274 (distinct)	0.40**	-0.41***	0.29***	-0.13***	-0.39**	-0.31**	-0.20*	0.30***	0.25***
Meta: partial		274 (distinct)	0.44**	-0.29***	0.31**	0.25***	-0.50***				-0.25*

Table S2: Cross correlation coefficients between transport costs metrics and density metrics. Linear population density is defined as the urban population divided by the square root of the municipal area. The cross correlations indicate that causally transport costs could be the primary driver of lower energy use (in transport), and population density is possibly only intermediate. Significance levels: ***p<0.01; **p<0.05; *p<0.1.

	Gasoline price vs. population density	Gasoline price vs. linear population density	gasoline price/GDP vs. population density
World Bank	0.52**	0.55**	
GEA		- 0.35***	
UITP			0.34**

3 Discussion of the emission and energy use elasticities

The regression analysis results presented in Table 1 (main text) demonstrates that urban energy use and/or GHG emissions change considerably with changes in the determinants. The results suggest that every 1% increase in GDP corresponds to a statistically significant increase of urban energy use and related emissions by 0.4%. Globally, the elasticity of CO_2 emissions with respect to GDP has an estimated value of 0.81 [10]. The difference between these two values suggests that direct urban energy use is considerably less sensitive to increases in economic activity then overall energy use of economies, such as production activities in non-urban areas and energy use induced by consumption. Another explanation is that cities use their energy more efficiently than rural areas making use of economies of scale in transport networks, infrastructures and housing [11]. Every 1% increase in population density corresponds statistically to a decrease of emissions of about 0.1–0.3%, possibly up to 0.6% in the case of transport energy use. The high elasticity of transport energy use, as reported by the UITP data set, needs to be put into context. First, the elasticity results are due to differences in urban form and transport energy use in different world regions [12]. Especially in North America, the elasticity of population density has been estimated to be considerably lower [13]. Second, population density may be a proxy for other characteristics of urban form such as land use mix, accessibility, or compactness [14]. In fact, when controlling for vehicle ownership, a likely driver of fuel consumption, the direct effect of population density on transport energy use is reduced [15]. The total effect of HDD on energy consumption and GHG emissions is considerable and highly significant and consistent across data sets. Every 1% increase in HDD corresponds to 0.3–0.5% increase in energy use or GHG emissions. Gasoline prices impact energy consumption and GHG emissions of cities to a high degree. Every 1% increase in gasoline price corresponds statistically to a decrease of emissions or energy use in the range of 0.4–0.8%. The elasticity of transport energy use with respect to gasoline prices (0.56) is consistent with the long-term fuel demand elasticity with respect to transport prices, observed in transport studies [16]. But total urban energy/GHG emissions change with gasoline price not only in the transportation sector. A possible explanation is that with higher transport prices, individuals will live closer to the city center, and that the higher density reduces energy demand for heating [17]. Another explanation is self-selection: choice about residential location is based on individual preferences. Thus, those who wish to behave environmentally and save energy move to cities that have features that enable alternative modes of transportation. While the second hypothesis cannot be excluded, the first hypothesis is supported by the observation that correlations between population density and transport costs metrics are significant, suggesting that transport cost is the primary factor, and population density is an inter-

mediate, secondary outcome (Tab. S2). Similarly, while household size is not significant in the regression model, household size correlates significantly ($p < 0.01$) with gasoline prices (normalized per GDP) in all three data sets. Hence fuel price reduces emissions from transport directly, but, when combined with reduced per capita floor space, also reduces emissions from housing (heating/electricity). CDD, population size, coastal city location, and household size are non significant in the regression analysis. Notably, this means that the efficiency effect of population size, relevant in the correlation analysis, can possibly be explained by other variables, such as population density.

4 Emission/Energy-driven top-down clustering

The reported analysis focused on clustering cities by their explanatory variables. To substantiate results, cities were also clustered by their emissions/energy use to group them into high, medium, and low emitters (Table S3). The city clusters can be characterized by underlying attributes that are statistically distinct for each city class (ANOVA test used for statistical significance; Table S3). High-emissions cities are consistently associated with high economic activity and low population density and to lesser degree with high HDD and low gasoline prices. Low-emissions cities display high population density and low economic activity and tentatively less HDD and higher gasoline prices. Table S3 demonstrates that GDP, population density, and to lesser degree HDD and gasoline prices are statistical predictors of emission/energy usage. In the two data sets with larger number of case studies, GEA and UITP, low emission/energy cities are systematically associated with low GDP, but this relationship is not significant between high/medium emitters and high/medium GDP. This suggests that among cities within an income above $10k various emission and energy usage trajectories are possible - independent of economic activity. Low emission cities are also correlated with population density in two of three databanks. But the inverse is only partially true: low density seems to be most relevant for transport-related energy use (UITP) and less so for overall energy consumption and emissions. Altogether, the top-down emission-ranking method reconfirms the results from bottom-up, driver-based, clustering.

Table S3: Emission/energy driven clustering of cities, and its statistically significant properties. Metrics that distinguish a category from both others are denoted in bold letters, those who only distinguish a category from one other are denoted with in normal letters.

	World Bank	GEA	UITP
High emitters	**high GDP**	med.-high GDP	med.-high GDP
	med.-low density	med.-high HDD	**low density**
	high HDD		med.-high HDD
	low gasoline price		
Medium emitters	med.-low GDP	med.-high GDP	med.-high GDP
	med.-low density	med.-high HDD	**med. density**
Low emitters	med.-low GDP	**low GDP**	**low GDP**
	high density	**low HDD**	**high density**
	low HDD		
	high gasoline price		

5 Regression with quadratic income term

Table S4 shows the results of including a squared term in GDP per capita in the relationship between energy consumption and its determinants for the WB, GEA, and UITP data. The estimated coefficient for the squared GDP per capita term is negative for GEA and UITP, but is statistically significant only for the latter. This could be interpreted as lending some support to the so-called *Environmental Kuznets curve hypothesis* that posits the existence of an inverted-U relationship between environmental impact and GDP, originally proposed by [18]. However, the estimated turning points for GEA and UITP data, above which increases in income result in emission reductions, would fall well outside the range of observed income values to be of practical significance (about 665,500 and 150,300 more than 10 and almost 3 times more than the maximum for GEA and UITP GDP per capita, respectively). These findings are also consistent with the tree regression and threshold estimation nonparametric approach that does not find, even for the higher income group cities, any subsample with a negatively sloped GDP per capita income term.

Table S4: Regression with Quadratic GDP Term

	Dataset:		
	WB	GEA	UITP
GDP pc	−3.700	1.358*	1.502***
	(5.366)	(0.704)	(0.468)
Density	−0.378**	−0.070***	−0.551***
	(0.151)	(0.026)	(0.050)
$HDD_{15.5}$	0.134**	0.065**	−0.019
	(0.063)	(0.025)	(0.019)
Gasoline	−0.621	−0.377***	−0.323***
	(0.440)	(0.127)	(0.078)
$(\text{GDP pc})^2$	0.195	−0.051	−0.063**
	(0.269)	(0.037)	(0.026)
Constant	21.238	−3.916	3.337
	(26.577)	(3.321)	(2.053)
Observations	24	223	64
R^2	0.724	0.360	0.892
Adjusted R^2	0.647	0.345	0.881

Notes: *p<0.1; **p<0.05; ***p<0.01.
Variables are in logs.
To interpret the impact of GDP per capita in the presence of nonlinear terms we need to take the partial derivative of consumption w.r.t GDP pc. Evaluated at the mean level of GDP pc, they are .27, .37, and .30, respectively for the WB, GEA, and UITP datasets.

6 Splitting and threshold estimation tree-based method

In this paper we use the recursive data partitioning algorithm developed by [19] known as GUIDE, which stands for Generalized, Unbiased, Interaction Detection and Estimation. GUIDE is an extension of the well known classification and regression trees (CART) algorithms developed by [20] that repeatedly splits the data into increasingly homogeneous groups by fitting constant models and defining a simple rule based on a single explanatory variable until it becomes infeasible to continue. At each split the available sample is partitioned into two groups, obeying different linear models, based on a single best predictor variable, the variable that minimizes the sum of squared residuals from regression over all possible splits for all available independent variables. It then applies the same splitting procedure on each of the subset areas separately. The output can be represented as a binary tree with branches and terminal nodes. The predicted value at each terminal node is the average at that node. The goal is to partition the data into homogeneous group whilst simultaneously preventing the tree form getting too large. Typically a large tree is "grown" first which is then reduced in size by a suitable "pruning" procedure. CART's recursive approach is particularly well suited when there is a complex interaction structure among the explanatory variables, such as dependencies that may be hierarchical, nonlinear, or of higher order in nature. CART can also deal with missing values. CART type models can be viewed as parsimonious strategies for a fully nonparametric estimation of a regression model. Regression-tree methods are known to be consistent in the sense that, under standard statistical assumptions, the predicted values converge to the unknown nonlinear regression function values pointwise [21]. GUIDE improves on its predecessors by minimizing potential biases in variable selection and interaction detection and by allowing to fit a linear model at each node. This approach has been shown to improve the prediction accuracy of the resulting tree and its interpretability [22, 23]. More importantly, this approach grounds our classification to the theoretical relationship between relevant variables.

Table S5 reports the regression results and city membership for each node in Fig. 2. Estimation was performed using Loh's GUIDE software, available at `http://www.stat.wisc.edu/~loh/guide.html` (last accessed September 2014).

Table S5: Tree Regression Node Results

	Node Subset:							
	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
GDP pc		0.894**			0.578	0.571***	0.368***	
		(0.120)			(0.189)	(0.112)	(0.098)	
HDD		−0.958			−0.240	−0.042	0.145***	
		(0.341)			(0.044)	(0.029)	(0.041)	
Density				0.319***	−0.466*	−0.120*	−0.134***	0.340**
				(0.107)	(0.065)	(0.071)	(0.023)	(0.143)
Gasoline		12.084**		4.615***	2.354			
		(1.559)		(1.453)	(0.672)			
Constant	3.694***	4.478	2.957***	2.209***	4.087	−0.019	0.547	2.400**
	(0.471)	(2.817)	(0.197)	(0.746)	(1.357)	(1.058)	(0.935)	(1.026)
Observations	5	6	5	24	6	36	127	14
R^2	0.000	0.985	0.000	0.479	0.985	0.448	0.410	0.319
Adjusted R^2	0.000	0.963	0.000	0.430	0.927	0.396	0.396	0.262

Notes: *p<0.1; **p<0.05; ***p<0.01.
Variables are selected using forward-and-backward stepwise regression in each node (for more details see the GUIDE documentation.
Node membership:

- Node 8
  - ASIA: Kunming, Nanning
  - MAF: King Sabata , Mangaung , Sedibeng
- Node 9
  - ASIA: Haerbin, Jilin, Lanzhou, Wulumqi, Xining
  - MAF: Sol Plaatje
- Node 10
  - ASIA: Ahmadabad, Bangalore, Haikou
  - MAF: Dar es Salaam, Nakuru
- Node 11
  - ASIA: Beijing, Changchun, Changsha, Hunan, Chengdu, Chongqing, Fuzhou, Guiyang, Hefei, Nanchang, Shanghai, Shijiazhuang, Taiyuan, Tianjin, Wuhan, Xi'an, Shaanxi, Yinchuan, Zhengzhou
  - MAF: Buffalo City , Ekurhuleni , eThekwini , Msunduzi , Nelson Mandela , Potchefstroom
  - REF: Kiev
- Node 12
  - ASIA: Dalian, Hohhot
  - MAF: Saldanha uMhlatuze OECD: Boulder, Sydney
- Node 13
  - ASIA: Bangkok, Guangzhou, Hangzhou, Hong Kong, Iskandar, Jinan, Shandong, Nanjing, Ningbo, Qingdao, Shenyang, Shenzhen, Singapore, Xiamen
  - LAC: Mexico City
  - MAF: Cape Town, Johannesburg, Tshwane
  - OECD90: Athens, Austin, Barcelona, Denver, Fort Collins, Los Angeles, Madrid, Melbourne, Minneapolis, New York, Portland, Seattle, Toronto
  - REF: Bucharest, Ljubljana, Moscow, Riga, Tallinn, Vilnius
- Node 14
  - LAC: Rio de Janiero, Sao Paulo
  - OECD90: Amsterdam, Barnsley, Bath, Bedfordshire, Belfast, Berkshire, Berlin, Birmingham, Blackburn, Bournemouth, Bradford, Bremen, Bridgend, Brighton and Hove, Bristol, Brussels, Buckinghamshire, Calderdale, Cambridgeshire, Cardiff, Central Valleys, Cheshire, Clackmannanshire, Copenhagen, Darlington, Derby, Derbyshire, Dublin, Dunbartonshire, Durham, East Cumbria, East Derbyshire, East Lothian and Midlothian, East Merseyside, East of Northern Ireland, East Riding of Yorkshire, East Sussex, Edinburgh, E N Ayrshire, Essex, Falkirk, Flintshire, Genva canton, Glasgow, Gloucestershire, Gwent Valleys, Halmstad, Halton and Warrington, Hamburg, Hampshire, Hartlepool, Herefordshire, Hertfordshire, Inverclyde, Istanbul, Kent, Kingston upon Hull, Kyoto, Lancashire, Leeds, Leicester, Leicestershire, Lincolnshire, Lisbon, Liverpool, London, Luton, Lyon, Manchester, Medway, Milan, Milton Keynes, Monmouthshire, N NE Lincolnshire, N Nottinghamshire, Norfolk, Northamptonshire, North Lanarkshire, North Yorkshire, Nottingham, Outer Belfast, Oxfordshire, Paris, Peterborough, Plymouth, Portsmouth, Rome, S Ayrshire, Sefton, Sheffield, Shropshire, Skane county, S Lanarkshire, Somerset, Southampton, Southend-on-Sea, South Nottinghamshire, Staffordshire, Stoke-on-Trent, Suffolk, Sunderland, Surrey, Swansea, Swindon, Telford, Thurrock, Tokyo, Turin, Tyneside, Vienna, Warwickshire, West Cumbria, West Lothian, West Sussex, Wiltshire, Wirral, Worcestershire, York, Zurich
  - REF: Belgrade, Bratislava, Budapest, Prague, Sofia, Zagreb
- Node 15
  - ASIA: Bangkok, Guangzhou, Hangzhou, Hong Kong, Iskandar, Jinan, Shandong, Nanjing, Ningbo, Qingdao, Shenyang, Shenzhen, Singapore, Xiamen
  - LAC: Mexico City
  - MAF: Cape Town, Johannesburg, Tshwane
  - OECD90: Gävle, Helsinki, Jönköping, Karlstad, Linköping, Norrköping, Örebro, Oslo, Stockholm county, Sundsvall, Umeå, Uppsala, Västra Götaland, Växjö
  - REF: Bucharest, Ljubljana, Moscow, Riga, Tallinn, Vilnius

7 Cross-validation analysis to select the size of the tree

Another concern is the problem of over- or under-fitting the data with nodes. Tree structure, including the number of nodes, and therefore the classification of cities is determined by the data rather than specified a priory. Specifically, we used an established cross-validation methodology to determine the optimal size of trees that minimizes miss-classification errors.

The methodology is called cost-complexity pruning, first introduced by [20] to determine the optimal size of trees that minimizes miss-classification errors. Following best practice [see, e.g., 24] we split the training set of cities into 10 roughly equally sized parts. We can then use 9 parts to grow the tree and test it on the tenth. This can be done in 10 ways, and we can average the results. This is known as (10-fold) cross-validation. Figure S2 show the results of the cross-validation analysis of selection of the size of the tree based on miss-classification rate. To minimize the impact of outliers and non-normality we used both the mean and median bootstrap square error estimates. The figure show that optimal number of terminal nodes is between 8 and 9. Estimation was also performed using Loh's GUIDE software, available at `http://www.stat.wisc.edu/~loh/guide.html` (last accessed September 2014).

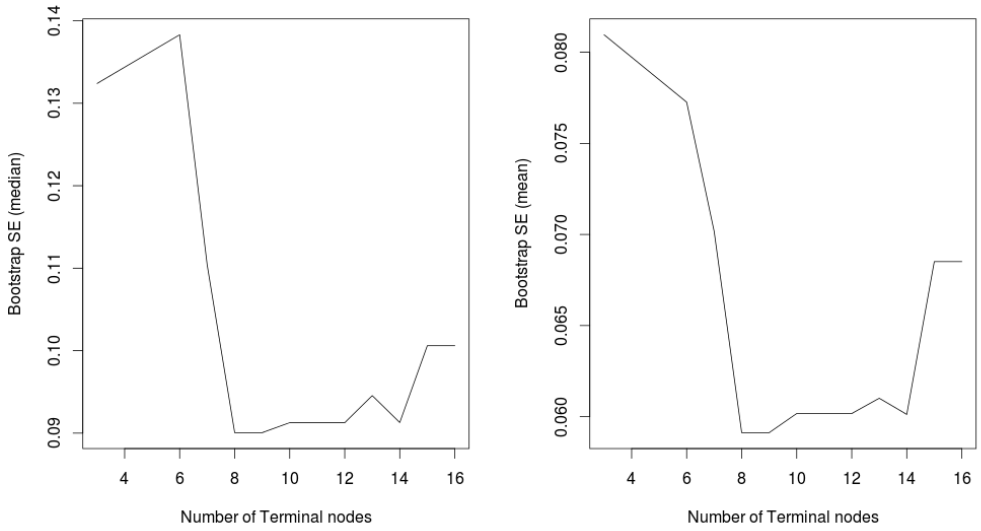


Figure S2: Plot of deviance (prediction error based on median and mean measured by the bootstrap squared difference between the observed and predicted values) versus the size (number of terminal nodes) of subtrees of the unpruned 16 terminal nodes tree. The results are based on pruning based on 10-fold cross-validation. Split values based on exhaustive search. Max number of split levels = 5. Minimum node size = 5. Number of SE's for pruned tree = 0.5

8 Confidence interval estimation for threshold values

One limitation of the tree regression approach is lack of asymptotic distribution theory useful for inference on splitting variables and split values [20]. [25] developed a threshold estimation testing procedure with accompanying distribution theory that addresses this issue. The procedure is widely used in economics and energy studies. We used the Hansen threshold regression approach to cross-validate our results and obtain confidence intervals for the main splits (compare with [22], [23]). The confidence intervals provide a measure of uncertainty in the classification of cities to specific types. In this section we briefly introduce the threshold estimation and testing procedure developed in [25] and [26] used to validate the tree regression results and to estimate confidence intervals for the main splits.

Let $\{y_i, \boldsymbol{x}_i, q_i\}$ be an observed sample, where $y_i,\ q_i \in \mathbb{R}$ and $\boldsymbol{x}_i = (1, x_{i2}, \cdots, x_{ik})^T$. The threshold variable q_i, which can be an element of $\boldsymbol{x}_i$, is assumed to have a continuous distribution. The threshold regression model

$$y_i = \boldsymbol{\vartheta}' \boldsymbol{x}_i + \boldsymbol{e}_i, \quad q_i \leq \tau \tag{S1}$$

$$y_i = \boldsymbol{\theta}' \boldsymbol{x}_i + \boldsymbol{e}_i, \quad q_i > \tau \tag{S2}$$

where $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2, \ldots, \vartheta_n)^T$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_n)^T$. After defining the dummy variable,

$$d_i(\gamma) = 1_{\{q_i \leq \tau\}},$$

the model (S1)-(S2), can be written as one equation

$$y_i = \boldsymbol{\theta}^T \boldsymbol{x}_i + \boldsymbol{\delta}^T \boldsymbol{x}_i d_i(\tau) + \boldsymbol{e}_i, \tag{S3}$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_n)^T$. Equation (S3) allows all parameters to differ across regimes. Keeping γ fixed, (S3) is linear in $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ and can be estimated by OLS. $\widehat{\gamma}$ can be defined as

$$\widehat{\gamma} = \underset{\tau \in \mathrm{G}_n}{\arg\min}\, S_n(\gamma).$$

where G_n is a suitably bounded set and S_n (concentrated) sum of squared error. [26] showed that, under some regularity conditions, the distribution of $\widehat{\gamma}$ is nonstandard but free of nuisance parameters.

To test the hypothesis $\mathrm{H}_0 : \gamma = \gamma_0$, a likelihood ratio approach can be employed with test statistic

$$LR_n = n \frac{S_n(\gamma) - S_n(\widehat{\gamma})}{S_n(\widehat{\gamma})} \tag{S4}$$

For large values of the statistic (S4) the null H_0 is rejected. [26] determines its asymptotic distribution.

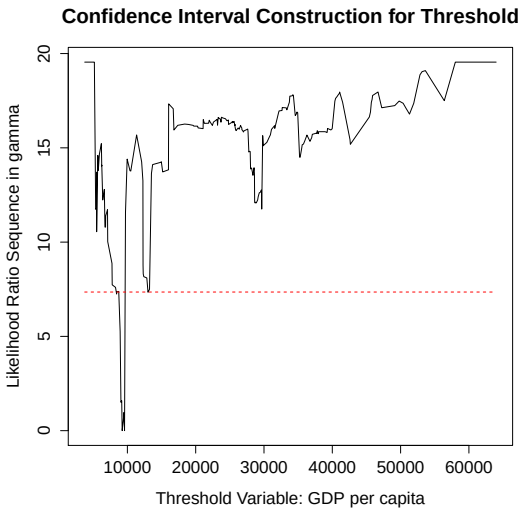
Confidence regions based on the likelihood ratio statistic can be obtained by inverting the likelihood ratio test of H_0: $\tau = \tau_0$. Denoting with c the relevant critical value for the distribution of the threshold, the confidence set is defined as

$$\widehat{\mathrm{T}} = \{\gamma | LR_n \leq c\}\,. \tag{S5}$$

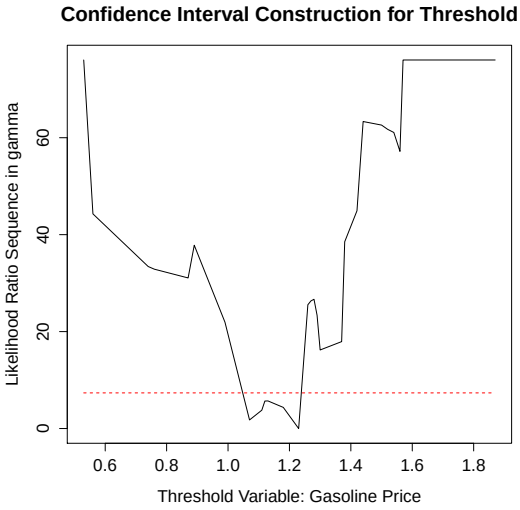
Hansen also provides the heteroskedasticity-robust asymptotic confidence set for γ, $\widehat{\mathrm{T}}^*$, based on a scaled version of the likelihood (denoted as LR^*), that are used in this paper. The main limitation of this approach is that it is limited to one threshold variable, one threshold value. See [26] and [23] for details.

The confidence intervals and their constructions presented in Figure 2 in the paper are reported in Figures S3a, S3b, S4a, S4b, S5a S5b, and S6a, for nodes 1, 2, 3, 4, 5, 6, and 7, respectively. Figure S6b shows the confidence interval construction for the GDP per capita split in UITP data presented in Figure S7. The asymptotic 95 % confidence set, $\widehat{\mathrm{T}}^*$ which is given in the graph by the levels where the $LR_n^*(\gamma)$ sequence crosses the dashed line. Whenever there is only one value below the dashed line, to avoid the problem caused by the low number of observations (for example in node 4 and 7), we provide a conservative estimate of the 95%CI by reporting the "bracketing" values of γ adjacent to the minimum. Estimation was performed using Hansen's code available at http://www.ssc.wisc.edu/~bhansen/progs/ecnmt_00.html (last accessed September 2014).

Figure S3: Confidence interval construction for nodes 1 and 3

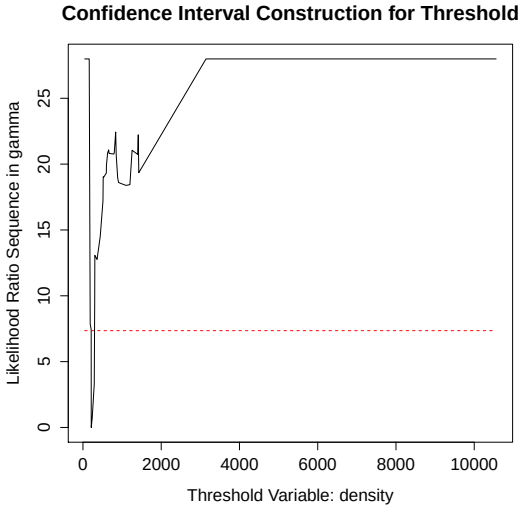


(a) Node 1. GDP sample split for the GEA data. The graph shows the heteroskedasticity-robust likelihood ratio sequence $LR_n^*(\gamma)$ against the threshold in natural log of GDP. The least square estimate of γ is the value that minimizes the curve, which occurs at about $\widehat{\gamma} = \$9,700$. The 95 % critical value of 7.35 is also plotted (dashed line). The asymptotic 95 % confidence set is $\widehat{\mathrm{T}}^* = [\$9,500, \$13,000]$, which in the graph is given by the levels where the $LR_n^*(\gamma)$ sequence crosses the dashed line. The result shows strong evidence for a GDP split confirming the tree regression results.

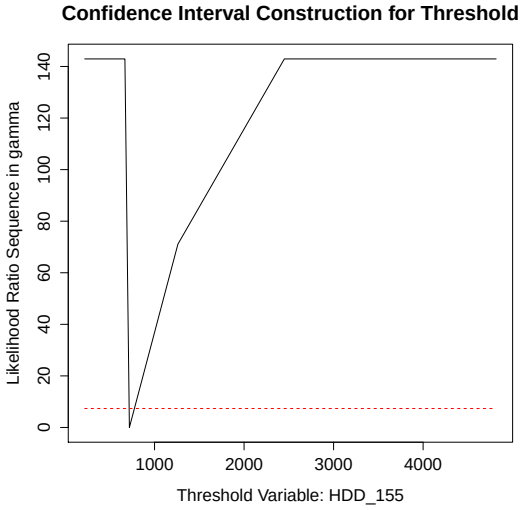


(b) Node 3. Gasoline price sample split for the GEA data. The graph shows the heteroskedasticity-robust likelihood ratio sequence $LR_n^*(\gamma)$ against the threshold in natural log of gasoline price. The least square estimate of γ is the value that minimizes the curve, which occurs at about $\widehat{\gamma} = 1.2\,\$/l$. The 95 % critical value of 7.35 is also plotted (dashed line). The asymptotic 95 % confidence set is $\widehat{\mathrm{T}}^* = [1.07\,\$/l, 1.23\,\$/l]$, which in the graph is given by the levels where the $LR_n^*(\gamma)$ sequence crosses the dashed line. The result shows strong evidence for a price split for "high" income cities confirming the tree regression results.

Figure S4: Confidence interval construction for nodes 2 and 4

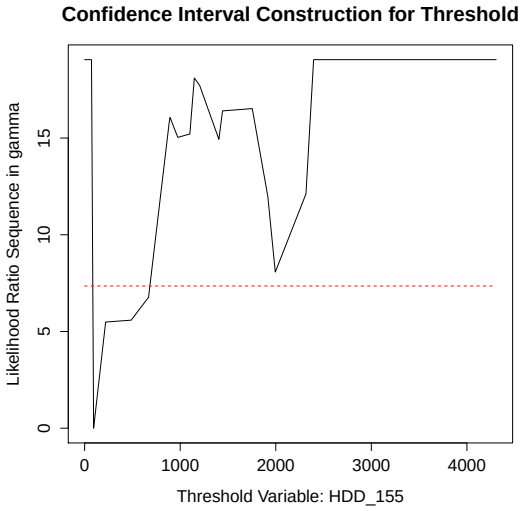


(a) Node 2. Density sample split for the GEA data. The graph shows the heteroskedasticity-robust likelihood ratio sequence $LR_n^*(\gamma)$ against the threshold in natural log of the density. The least square estimate of γ is the value that minimizes the curve, which occurs at about $\widehat{\gamma} = 210\,pop/km^2$. The 95% critical value of 7.35 is also plotted (dashed line). The asymptotic 95% confidence set is $\widehat{\mathrm{T}}^* = [200\,pop/km^2, 300\,pop/km^2]$, which in the graph is given by the levels where the $LR_n^*(\gamma)$ sequence crosses the dashed line. The result shows strong evidence for a density split for "low" income cities confirming the tree regression results.

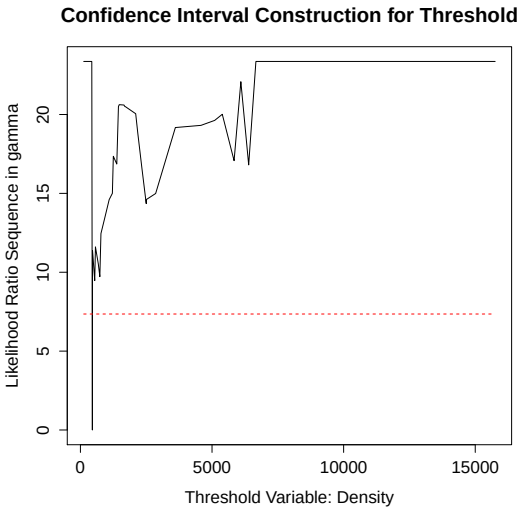


(b) Node 4. HDD sample split for the GEA data. The graph shows the heteroskedasticity-robust likelihood ratio sequence $LR_n^*(\gamma)$ against the threshold in natural log of HDD. The least square estimate of γ is the value that minimizes the curve, which occurs at $\widehat{\gamma} = 94$ HDD. The 95% critical value of 7.35 is also plotted (dashed line). The asymptotic 95% confidence set is $\widehat{\mathrm{T}}^* = [70, 670]$ HDD, because of the sparcity of the data, is given by the values of γ adjacent to the minimum. See detail in the text. which in the graph is given by the levels where the $LR_n^*(\gamma)$ sequence crosses the dashed line. The result shows strong evidence for an HDD split for lower density node 2 cities.

Figure S5: Confidence interval construction for nodes 5 and 6

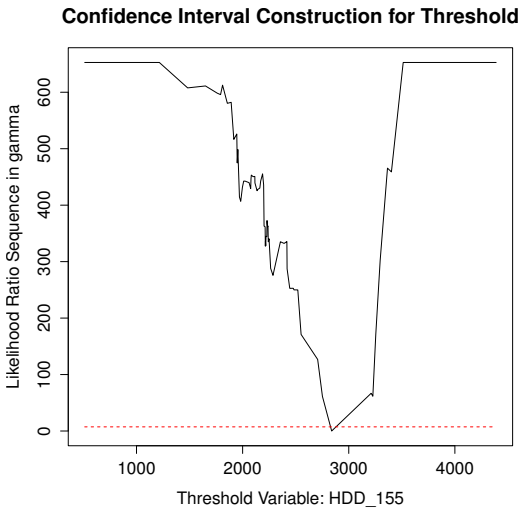


(a) Node 5. HDD sample split for the GEA data. The graph shows the heteroskedasticity-robust likelihood ratio sequence $LR_n^*(\gamma)$ against the threshold in natural log of HDD. The least square estimate of γ is the value that minimizes the curve, which occurs at $\widehat{\gamma} = 1261$ HDD. The 95 % critical value of 7.35 is also plotted (dashed line). The asymptotic 95 % confidence set is $\widehat{\mathrm{T}}^* = [719\,\mathrm{HDD}, 2450\,\mathrm{HDD}]$, which in the graph is given by the levels where the $LR_n^*(\gamma)$ sequence crosses the dashed line. The result shows strong evidence for an HDD split for higher density node 3 cities.

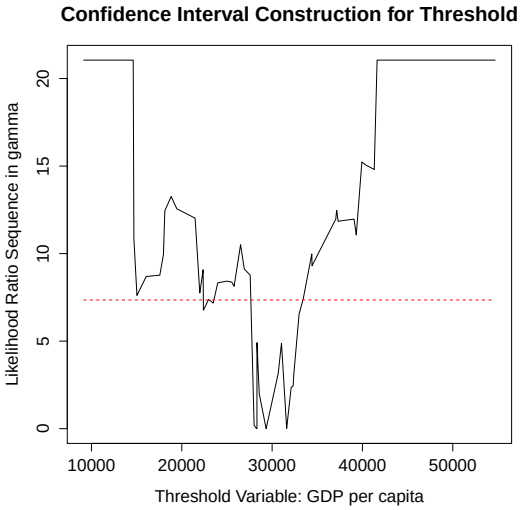


(b) Node 6. Density sample split for the GEA data. The graph shows the heteroskedasticity-robust likelihood ratio sequence $LR_n^*(\gamma)$ against the threshold in natural log of density. The least square estimate of γ is the value that minimizes the curve, which occurs at $\widehat{\gamma} = 454\,pop/km^2$. The 95 % critical value of 7.35 is also plotted (dashed line). The asymptotic 95 % confidence set is $\widehat{\mathrm{T}}^* = [453\,pop/km^2, 460\,pop/km^2]$, which in the graph is given by the levels where the $LR_n^*(\gamma)$ sequence crosses the dashed line. The result shows strong evidence for an density split.

Figure S6: Confidence interval construction for nodes 7 and the UITP GDP split



(a) Node 7. HDD sample split for the GEA data The graph shows the heteroskedasticity-robust likelihood ratio sequence $LR_n^*(\gamma)$ against the threshold in natural log of HDD. The least square estimate of γ is the value that minimizes the curve, which occurs at $\widehat{\gamma} = 2840$ HDD. The 95 % critical value of 7.35 is also plotted (dashed line). The asymptotic 95% confidence set is $\widehat{\mathrm{T}}^* = [2752, 3211]$ HDD, because of the sparcity of the data, is given by the values of γ adjacent to the minimum. See detail in the text.



(b) GDP sample split for the UITP data. The graph shows the heteroskedasticity-robust likelihood ratio sequence $LR_n^*(\gamma)$ against the threshold in natural log of GDP pc.. The least square estimate of γ is the value that minimizes the curve, which occurs at about $\widehat{\gamma} = \$29{,}300$. The 95% critical value of 7.35 is also plotted (dashed line). The asymptotic 95% confidence set is $\widehat{\mathrm{T}}^* = [\$22{,}400, \$33{,}000]$, which in the graph is given by the levels income where the $LR_n^*(\gamma)$ sequence crosses the dashed line. The result shows strong evidence for a GDP split.

9 Peak urban travel

We applied the GUIDE algorithm (see below) to the UITP data and found a threshold regression value at GDP/cap at \$29,300 with the confidence interval (CI) ranging at the 95% confidence level from \$22,400 until \$33,000 (detail on the construction of the CI are provided in Section 8 and Figure S6b). This is visualized in Figure S7. Transport energy use decreases under certain conditions with increasing economic activity within the high economic activity (affluent) segment. Transport energy use decreases with GDP for affluent cities in OECD90 countries below 2 million inhabitants. Cities with larger population density (size of circles), and with higher gasoline price tend to be associated with lower energy consumption. Both x-axis and y-axis display logarithmic scales. The elasticity of energy consumption at the lowest GDP per capita of \$400 is 0.9; the elasticity at average GDP per capita of \$21,400 is -0.5, and the elasticity at highest GDP per capita of \$55,000 is -0.9. The group of cities smaller than 2 million inhabitants and from OECD90 countries, as defined in the UITP data set [4], display significant decrease of urban transport energy use with income at $p < 0.01$. The significance increases to $p < 0.05$ if the outlier, Denver, is excluded from this analysis. As discussed in the main body text, this effect is closely associated with a continental distribution of cities.

10 Calculating the urbanization wedge

First, energy use of cities in the five world regions was scaled with expected population growth to 2050 (Tab. S4). Additional energy use was assumed to come from growth in economic activity per capita. In the median scenario (see below for uncertainty analysis), the GDP growth in the world regions was estimated to follow the B1 SRES scenario of the IPCC [27]. The GDP growth was translated by the GDP/cap elasticity (0.66 as long as GDP/cap$<$ \$9,700; 0.33 otherwise; calculated by regressions on node 2 and node 3 statistics in Figure 2), resulting in a global energy consumption of 731 EJ in 2050. Second, gasoline price was assumed to grow to \$1.6 in terms of 2005 \$ worldwide. This would translate into little change in some OECD countries but tremendous change in countries that currently subsidize fuels. The gasoline price elasticity (0.46; for regression on node 3, valid for GDP/cap$>$ \$9,700) was applied then to the difference in gasoline prices between 2005 and 2050. Finally, the population density is expected to grow proportionally to half of the population growth in each world region. In other words, already urbanized world regions gasoline price increase, this results into global urban energy consumption are expected to display relatively little potential for densification, a conservative assumption. The population density elasticity (0.09; for regression on node 3 cities, valid for GDP/cap$>$ \$9,700) was applied on this additional population density. Combined with the change due to of 540 EJ in 2050. To represent the underlying parameter uncertainty

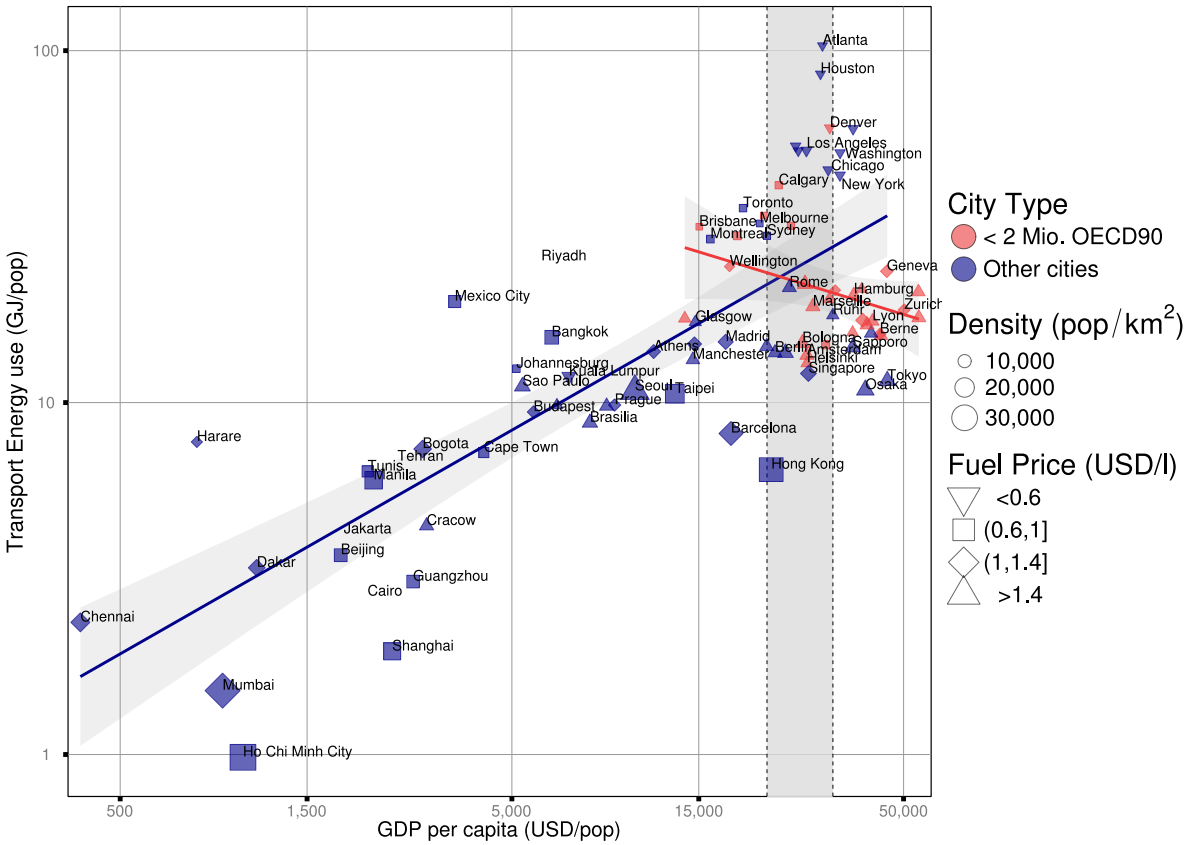


Figure S7: Urban peak travel (UITP data). Transport energy use decreases under certain conditions with increasing income within the high-income segment. Transport energy use decreases with GDP for high-income cities in developing countries below 2 million inhabitants. Cities with larger population density (size of circles), and with higher gasoline price tend to be associated with lower energy consumption.

(Fig. 5B), we performed a Monte Carlo simulation on sensitive parameters. Specifically, we draw randomly from a uniform distribution of GDP growth rates (globally: 1.7%–3.7% but weighted across world regions, taken from the SRES scenarios (18)); and from Gaussian distributions of the elasticities derived from the threshold regression (GDP: 0.67 (mean) $\pm$0.26 (standard deviation) if GDP/cap $<$ \$9,700; 0.30 ± 0.07 else; gasoline price: 0.46 ± 0.11; population density: 0.09 ± 0.03).

References

[1] Christopher Kennedy, Julia Steinberger, Barrie Gasson, Yvonne Hansen, Timothy Hillman, Miroslav Havranek, Diane Pataki, Aumnad Phdungsilp, Anu Ramaswami, and Gara Villalba Mendez. Greenhouse Gas Emissions from Global Cities. *Environ. Sci. Technol.*, 43(19):7297–7302, 2009.

[2] Christopher A Kennedy, Anu Ramaswami, Sebastian Carney, and Shobhakar Dhakal. Greenhouse gas emission baselines for global cities and metropolitan regions. *Cities and climate change: Responding to an urgent agenda*, pages 15–54, 2011.

[3] Arnulf Grübler, Xuemei Bai, Thomas Buettner, Shobhakar Dhakal, David Fisk, T Ichinose, James Keirstead, G Sammer, David Satterthwaite, Niels Schulz, Nilay Shah, Julia Steinberger, and Helga Weisz. Urban Energy Systems. In *Global Energy Assessment: Toward a Sustainable Future*. IIASA, Austria and Cambridge University Press, Cambridge / New York, 2012.

[4] J Kenworthy and F Laube. THE MILLENNIUM CITIES DATABASE FOR SUSTAINABLE TRANSPORT. DATABASE. 2001.

[5] U K PricewaterhouseCoopers. *Economic Outlook November 2009*. London, 2009.

[6] European Commission. Urban Audit: methodological handbook. *Luxemburg: Office for Official Publications of the European Commission*, 2004.

[7] Bizee. A free worldwide data calculation site on Heating & Cooling Degree Days (www.degreedays.net), last accessed September 2014.

[8] Sebastian Ebert. *International fuel prices 2009*. Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ), 2009.

[9] MasterCard Worldwide. Worldwide centers of commerce index. *available online at:* HYPERLINK" http://www. mastercardworldwide. com/insights"** www. mastercardworldwide. com/insights*, 2008.

[10] Edgar G Hertwich and Glen P Peters. Carbon Footprint of Nations: A Global, Trade-Linked Analysis. *Environ. Sci. Technol.*, 43(16):6414–6420, 2009.

[11] Luís M A Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17):7301–7306, 2007.

[12] A H Baur, M Thess, B Kleinschmit, and F Creutzig. Urban Climate Change Mitigation in Europe - Looking At and Beyond the Role of Population Density. *Journal of Urban Planning and Development*, 10.1061, 2013.

[13] R Ewing. Growing Cooler:The Evidence on Urban Development and Climate Change, 2007.

[14] Reid Ewing and Robert Cervero. Travel and the Built Environment – A Meta-Analysis. *Journal of the American Planning Association*, 76(3):265–294, 2010.

[15] Niovi Karathodorou, Daniel J Graham, and Robert B Noland. Estimating the effect of urban density on fuel demand. *Energy Economics*, 32(1):86–92, 2010.

[16] DANIEL J Graham and STEPHEN Glaister. Road Traffic Demand Elasticity Estimates: A Review. *Transport Reviews*, 24(3):261–274, 2004.

[17] Reid Ewing and Fang Rong. The impact of urban form on U.S. residential energy use. *Housing Policy Debate*, 19(1):1–30, 2008.

[18] G. M. Grossman and A. B. Krueger. Economic growth and the environment. *Quaterly Journal of Economics*, 110(2):353–377, 1995.

[19] W. Loh. Regression trees with unbiased variable selection and interaction detection. *Statistica Sinica*, 12:361–386, 2002.

[20] L Breiman, J H Friedman, R A Olshen, and C J Stone. *Classification and Regression Trees*. Wadsworth, Belmont, 1984.

[21] Probal Chaudhuri, Min-ching Huang, Wei-yin Loh, and Ruji Yao. Piecewise-polynomial regression trees. *Statistica Sinica*, 4:143–167, 1994.

[22] W. Loh. Improving the precision of classification trees. *The Annals of Applied Statistics*, 4:1233–1830, 2009.

[23] Chih Ming Tan. No one true path: uncovering the interplay between geography, institutions, and fractionalization in economic development. *Journal of Applied Econometrics*, 25(7):1100–1127, 2010.

[24] L Clark and D Pregibon. Statistical Models in S, chapter Tree-Based models. *Wadsworth International Group*, 1992.

[25] Bruce E Hansen. Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis. *Econometrica*, 64(2):413–30, 1996.

[26] Bruce E. Hansen. Sample splitting and threshold estimation. *Econometrica*, 68(3):575–604, 2000.

[27] Nebojsa Nakicenovic and Robert Swart. Special report on emissions scenarios. *Special Report on Emissions Scenarios, Edited by Nebojsa Nakicenovic and Robert Swart, pp. 612. ISBN 0521804930. Cambridge, UK: Cambridge University Press, July 2000.*, 1, 2000.