Global typology of urban energy use and potentials for an urbanization mitigation wedge

Felix Creutzig,b,c,1 Giovanni Baiocchi,2 Robert Bierkandtd, Peter-Paul Pichler4 and Karen C. Seto* 

*Mercator Research Institute on Global Commons and Climate Change, 10829 Berlin, Germany; bTechnische Universität Berlin, 10623 Berlin, Germany; cDepartment of Geographical Sciences, University of Maryland, College Park, MD 20742; dPotsdam Institute for Climate Impact Research, 14473 Potsdam, Germany; and *Yale School of Forestry & Environmental Studies, New Haven, CT 06511

Edited by Sangwon Suh, University of California, Santa Barbara, CA, and accepted by the Editorial Board December 4, 2014 (received for review August 17, 2013)

The aggregate potential for urban mitigation of global climate change is insufficiently understood. Our analysis, using a dataset of 274 cities representing all city sizes and regions worldwide, demonstrates that economic activity, transport costs, geographic factors, and urban form explain 37% of urban direct energy use and 88% of urban transport energy use. If current trends in urban expansion continue, urban energy use will increase more than threefold, from 240 EJ in 2005 to 730 EJ in 2050. Our model shows that urban planning and transport policies can limit the future increase in urban energy use to 540 EJ in 2050 and contribute to mitigating climate change. However, effective policies for reducing urban greenhouse gas emissions differ with city type. The results show that, for affluent and mature cities, higher gasoline prices combined with compact urban form can result in savings in both residential and transport energy use. In contrast, for developing-country cities with emerging or nascent infrastructures, compact urban form, and transport planning can encourage higher population densities and subsequently avoid lock-in of high carbon emission patterns for travel. The results underscore a significant potential urbanization wedge for reducing energy use in rapidly urbanizing Asia, Africa, and the Middle East.

Significance

Many case studies of specific cities have investigated factors that contribute to urban energy use and greenhouse-gas emissions. The analysis in this study is based on data from 274 cities and three global datasets and provides a typology of urban attributes of energy use. The results highlight that appropriate policies addressing urban climate change mitigation differ with type of city. A global urbanization wedge, corresponding in particular to energy-efficient urbanization in Asia, might reduce urban energy use by more than 25%, compared with a business-as-usual scenario.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. S.S. is a guest editor invited by the Editorial Board.

1To whom correspondence should be addressed. Email: creutzig@mcc-berlin.net.

This article contains supporting information online at www.pnas.orglookup/suppl?doi:10.1073/pnas.1315545112/DCSupplemental.
direct final energy use of cities, including electricity consumption and heating in building and energy use from urban transport from the GEA. We used information on final energy use in the urban transport sector provided by the UITP, and the WB data on GHG emissions in cities, comprising energy use, waste, and industrial processes but excluding marine and aviation emissions. We tested the robustness, stability, and statistical relationships across the datasets using advanced statistical and data-mining methods (Methods).

Although fuel composition often changes with urbanization (17), fuel composition and consumption-based emissions were not considered in this study. Limiting our scope to energy end-use within cities and direct emissions enabled this study to identify urbanization-relevant attributes that are not often included in technology-focused studies of climate stabilization. We did not analyze cross-boundary contributions to GHG emissions and consumption-based carbon footprints, studied elsewhere (6, 13). In total, our dataset (Dataset S1) includes 274 cities from 60 countries with a combined population of 775 million, or 21% of the global urban population (SI Appendix, Section 1; all data provided in Dataset S1). The cities in the study include all 21 megacities (cities with more than 10 million inhabitants) and are distributed across a range of population sizes, with the smallest city of 55,000.

We calculate the rank size of the entire sample cities to examine whether the 274 cities in the study are statistically representative of cities worldwide in terms of their size (18). The frequency of cities of different population sizes and their rank, where rank is determined by the frequency of occurrence, are connected through a power-law function $P(i) \sim i^{-\alpha}$ with the exponent $\alpha$ close to unity, giving rise to the so-called Zipf Law (19). Sometimes called the rank-versus-frequency rule, it is a mathematical formulation of a long-observed phenomenon: the city with the largest population is about twice the size of the second largest and so on. Our results show that it scales log-log-linearly with population size (SI Appendix, Fig. S1). The slope between log city rank and log city size is $-1.07$, and the 95% confidence interval (CI) is between $-1.21$ and $-0.93$, including the value $-1$. This result confirms that the city sizes in this study are broadly representative of the global system of cities.

Results

Our analysis (Fig. 1 and SI Appendix, Section 2) shows that gasoline price and population density correlate most strongly with transport energy use and GHG emissions, followed by economic activity. In contrast, the effect from economic activity dominates final energy consumption and is followed in importance by climatic variables [heating degree days (HDDs)], household size, and urbanization rate. The analysis shows that, across all datasets, economic activity, population density, and gasoline price are the most important factors in GHG emissions, but population size, household size, urbanization level, and an index of commercial centrality are also significant (SI Appendix, Table S1). However, out of the last four, only the commercial centrality index remains significant in the partial regression when we control for the other variables (SI Appendix, Table S1). Surprisingly, energy use decreases with an increase in cooling degree days (CDDs) (Fig. 1). This correlation is possibly an indirect effect of the concurrent reduction of HDD. A partial correlation, controlling inter alia for HDD, demonstrates a positive impact on energy use by increase in CDD (SI Appendix, Table S1). Coastal locations do not correlate significantly with either energy use or GHG emissions.

The regression models (Methods) also show that four factors explain at least one-third of the variance in urban energy end-use and GHG emissions. According to the regression results, transport energy use is driven by fuel price, population density, and economic activity (Table 1). Together, economic (economic activity, gasoline prices), structural (population density), and geographical (HDD, but not CDD and coastal city location) variables explain an important fraction of the energy use variability of cities (adjusted $R^2$: WB, 0.70; GEA, 0.35; UITP, 0.88) (Table 1). In general, economic factors (economic activity and gasoline prices) are more correlated with energy use and GHG emissions than structural variables (population density), whereas geographic variables (HDD) are highly significant but induce less marginal change in energy use (Table 1, substantiated by energy-driven top-down clustering; see SI Appendix, Sections 3 and 4). We also tested for nonlinear gross domestic product (GDP) terms, which were significant only for the UITP data (SI Appendix, Section 5 and Table S4, and discussion below).

Importantly, urban energy use is less elastic to changes in economic activity than observed in studies using national data (20). This observation may be a reflection of economies of scale in urban infrastructures and/or the relocation of energy-intensive urban production activities to rural areas. The elasticity of energy use with respect to fuel prices is significant, even in datasets of total direct urban energy use and emissions. This finding indicates that fuel prices influence energy use not only in transport, but possibly also in residential energy use. A plausible explanation for why fuel prices influence residential energy use is that, with higher transport prices, individuals will live closer to the city center and that the higher density reduces energy demand for heating (11) (SI Appendix, Section 3 and Table S2).

Urban Energy Use Typologies. Beyond the aggregate statistics that are reported above, contextual factors such as development stage and historical development as well as the interaction between various attributes, might be equally important for explaining emission patterns across cities. To examine this hypothesis, we developed a typology of cities in the GEA dataset according to the combination of their emission attributes using endogenous
threshold estimation and testing procedures in the regression context (GUIDE algorithm, see Methods; for cross-validation and confidence interval estimation, see SI Appendix, Sections 6–8).

The analysis showed that the relationship between energy and its determinants varies across cities. Our analysis identified eight types of cities which are characterized by a combination of GDP per capita, population density, gasoline price, and HDD. Affluence is the most important thresholding variable at the top level (Fig. 2). Among affluent cities, those with high or medium gasoline prices and high population density have lowest emissions (Fig. 2). At the second level, threshold values in both population density and gasoline prices separate types of cities (Fig. 2). At the third level, heating degree days (HDDs) and population density further split cities into different types (Fig. 2).

Cities with GDP less than 10,000 USD per capita (19% of all cities analyzed) show nearly three times lower energy use than those above this threshold (Figs. 2 and 3). Among less affluent cities (<10,000 USD per capita), the city type with the highest population densities (4,600 population per km²) and lowest HDD (30 in average) show the lowest energy use (~20 GJ per capita) (Fig. 3). More affluent cities (>10,000 USD per capita), in contrast, are clustered by both gasoline prices and population density. Among these affluent cities, those with gasoline prices above 1.2 USD/L and less than 3,000 HDD have relatively low density. Among these affluent cities, those with gasoline prices in contrast, are clustered by both gasoline prices and population density further split cities into different types (Fig. 2).

The relative importance of each individual factor (GDP per capita, population density, gasoline prices, HDD) changes with type of city (SI Appendix, Table S5).

Peak Urban Transport Energy Use. Our analysis of the UITP dataset corroborates the analysis with the GEA data and shows that energy use saturates with increasing economic activity, specifically for urban transport (SI Appendix, Section 9 and Fig. S2). Energy consumption for urban transport increases with GDP at low GDP levels but decreases with GDP at high GDP levels (threshold regression, 29,300 USD; CI, 22,400–33,000 USD). Specifically, cities in developed countries with GDP per capita over 13,500 USD and with fewer than two million inhabitants show a slight decrease in transport energy use with increasing GDP (Pearson correlation coefficient r = −0.35, P = 0.06). All other cities show strong growth in transport energy use with GDP per capita (r = 0.65, P < 0.01). This finding is evidence that,

### Table 1. Emission elasticities

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Dependent variable</th>
<th>No. of cities</th>
<th>GDP per capita</th>
<th>Pop. density</th>
<th>HDD</th>
<th>Fuel price</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>WB</td>
<td>GHGe per capita</td>
<td>26</td>
<td>0.18</td>
<td>−0.38**</td>
<td>0.13**</td>
<td>−0.76*</td>
<td>0.70</td>
</tr>
<tr>
<td>GEA</td>
<td>FE per capita</td>
<td>225</td>
<td>0.39***</td>
<td>−0.07***</td>
<td>0.07***</td>
<td>−0.37***</td>
<td>0.35</td>
</tr>
<tr>
<td>UITP</td>
<td>UT FE per capita</td>
<td>87</td>
<td>0.45***</td>
<td>−0.42***</td>
<td>0.02</td>
<td>−0.55***</td>
<td>0.88</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td>0.39***</td>
<td>−0.28**</td>
<td>0.13*</td>
<td>−0.55***</td>
<td></td>
</tr>
</tbody>
</table>

Fe, final energy; GHGe, greenhouse gas emissions; UT FE, urban transport final energy. Significance levels: ***P < 0.01; **P < 0.05; *P < 0.1.

The third-level thresholds in this clustering analysis show a surprising relevance of HDD. The linear regression model demonstrates a significant but weak influence of HDD. In contrast, the nonlinear threshold regression exhibits that HDD becomes a highly relevant variable, once affluence, transport prices, and urban form variables are controlled for. Among less affluent cities, energy consumption differs by a factor of three between cities in warmer (type 10) or cooler (type 11) climatic zones, and for more affluent cities by a factor of 1.5 (type 14 versus type 15) (Fig. 2). The less affluent types 8 and 9 display similar energy consumption at notably different HDD levels, but the lower level of HDD in type 8 is compensated by 10% lower gasoline costs (Fig. 2).

The analysis showed that the relationship between energy and its determinants varies across cities. Our analysis identified eight types of cities which are characterized by a combination of GDP per capita, population density, gasoline price, and HDD. Affluence is the most important thresholding variable at the top level (Fig. 2). Among affluent cities, those with high or medium gasoline prices and high population density have lowest emissions (Fig. 2). At the second level, threshold values in both population density and gasoline prices separate types of cities (Fig. 2). At the third level, heating degree days (HDDs) and population density further split cities into different types (Fig. 2).
at higher levels of income, urban transport decouples from GDP per capita, similar to what has been observed on the national level in OECD (Organisation for Economic Co-operation and Development) countries (21). Affluent cities with GDP per capita above 13,500 USD and population size below 2 million inhabitants, mostly European cities, have higher gasoline prices (P < 0.01) but lower population density than the other cities (P < 0.01). This result suggests that increases in economic activity increase the demand for urban land and thus reduce population density, but that an increase in gasoline taxes can mediate and even counterbalance this pattern.

An Urbanization Wedge. Our results point to the potential of urbanization to save energy and reduce associated GHG emissions. This raises the question of the potential magnitude of the global urbanization wedge between 2005 and 2050. In 2005, the global urban population of ~3.2 billion consumed about 240 EJ of energy at end-use (SI Appendix, Section 10) (22). By 2050, the urban population is expected to double to around 6.4 billion (2). During this same period, the average GDP per capita could plausibly increase by 80% in OECD countries and 390% in Asia (23). Accounting for population growth and considering the economic activity elasticity in low and affluent types of cities, total energy consumption in cities worldwide could increase to around 730 EJ (Fig. 4A and SI Appendix, Section 10). This estimate omits potential interventions by urban planning or gasoline taxes, and assumes stable oil prices. If urban planning and fuel taxes are used, total urban energy consumption could increase to about 540 EJ (Fig. 4A), assuming a universal increase of gasoline prices to 1.6 USD/L (deflated to 2005 dollars; approximately reflecting the existing level of gasoline prices in European countries) (24) and assuming urban planning policies that support higher population densities, mixed-use development, and accessibility (Methods). More precisely, we model population density to increase half as fast as population growth. That is, when the total population of a world region increases by 10% between 2005 and 2050, then urban planning allows urban population density to increase by 5%. The mitigation potential is greatest in rapidly growing cities and in cities with low gasoline prices. The total urbanization wedge is about 180 EJ. More than half of this urbanization wedge is in Asia (57%), and nearly one-third (29%) is in Africa and the Middle East. In contrast, the OECD90 (OECD countries in 1990) possesses relatively low potential for reducing energy use in cities (6%), mainly because OECD90 cities are mature, built-up environments with established infrastructure and associated locked-in behavior and energy consumption patterns.

The uncertainty underlying these scenarios is considerable. Structural uncertainty is based on uncertainties in economic growth, how urban energy use changes with economic growth, fuel prices, and population density changes. We rely on the spread of the SRES (Special Report on Emissions Scenarios) (24) and assuming urban planning policies confidence intervals obtained in the threshold regressions to estimate overall energy use in world regions in the business-as-usual and policy scenarios and report the uncertainty by a Monte Carlo simulation (Fig. 4B; for more details, see SI Appendix, Section 10). Despite the uncertainties, the scenarios illustrate the enormous potential for a mitigation wedge in urbanizing Asia, Africa, and the Middle East. The uncertainties change the magnitude of the potential, but the underlying reasons for the mitigation wedge remain. It is in those places where infrastructure is still nascent that there is the greatest mitigation potential.

Discussion

This study points to a considerable but differentiated potential for an urban mitigation wedge. Continued urban population growth and associated development of urban areas worldwide combined with projected increases in GDP per capita could lead to a tripling of urban energy use from 2005 to 2050. However, the cities of tomorrow develop spatially, especially their urban form, will lack in patterns of energy consumption for decades to come. Recent forecasts suggest that the global urban footprint will triple between 2000 and 2030, an area of 1.2
million km², or equal to the size of South Africa (3). Thus, mitigation interventions related to urban form have the highest potential during early phases of urban development. This window of opportunity exists especially for low-emissions cities in Asia, the Middle East, and Africa where urbanization and associated rises in income could lead to high increases in urban energy use if current trends continue. However, demand-side policies, such as increased gasoline prices, encouraging compact and accessible urban forms, along with idiosyncratic urban design options, can also reduce urban energy use in developed cities (25, 26). Backed by increases in gasoline prices, urban form modifications could reduce global energy use in cities by 26% or 190 EJ, constituting a notable and possibly low-cost or negative-cost urbanization wedge for climate change mitigation.

To realize this urbanization wedge, different types of cities require different mitigation strategies. Currently thousands of cities worldwide are developing local climate action plans, but their aggregate impact on emissions is uncertain (1). This uncertainty is in part because of low accountability and lack of baseline data on urban emissions, but also because the strategies adopted may not be the most effective at lowering emissions for a particular type of city. Although many urban mitigation strategies have important local co-benefits, any measurable impact on emissions will require adopting strategies that target the main sources of emissions. If countries with fuel prices below 1.2 USD/L increase this price to 1.6 USD/L, they would enable a market-based transition toward more energy-efficient cities. Similarly, urban-planning policies, including mixed-used design and high connectivity and accessibility, which are themselves closely related to population density, could be supportive in establishing long-term energy savings. In addition, those city types with high HDD could reduce emissions by enforcing stricter building codes and retrofitting strategies. These results are not only highly policy-relevant. In addition, the results provide a promising base for integrated assessment models that investigate the interaction between urbanization and technological decarbonization options in global scenarios.

This study provides for the first time, to our knowledge, robust (taking into account cross-data heterogeneity), statistically meaningful observations on a globally representative set of cities. The large sample of 274 cities validates some previously observed patterns while simultaneously providing new, statistically meaningful results (Fig. 1 and SI Appendix, Figs. S1–S7 and Tables S1–S5) and statistically and quantitatively significant elasticities (Table 1). High-emissions cities display in aggregate, high economic activity and low population density, low fuel prices, and high HDD whereas low-emissions cities have low economic activity, high population densities, and high fuel prices.

Fuel price as a potential driver of urban GHG emissions deserves particular attention for policy purposes. Although the importance of fuel prices as drivers of GHG emissions has been generally widely recognized (25), its specific importance for urban energy use and GHG emissions has not yet been systematically specified in the literature on cities and climate change. In the urban economics literature, gasoline prices and other transport costs have long been known to influence urban transport distances and modal choice, urban form, and population density (27, 28). Gasoline price is likely to influence GHG emissions directly and indirectly. Directly, in the long run, higher gasoline prices could induce a shift in private vehicle ownership but also could change patterns of where people live and work (30), travel behavior, and electricity/heating demand via modified floor space (11).

Our analysis focuses on the direct energy use in cities, including also the energy use from economic production activities. A number of studies have emphasized the importance of production activities for the urban carbon footprint (6, 13, 31). As a crude proxy, the commerce index correlation indicates that economic commerce activity plays a role also in our global sample of cities. In fact, production activity is likely to explain a significant part of the variability left unexplained by our study. For example, some of the outlier cities in Fig. 3 displaying very high energy use are small cities with oil refining or an allied industry as a dominating business. This hypothesis is supported by an analysis of the United Kingdom demonstrating that territorial emissions, including production-based carbon footprints of human settlements, are highly variable (13). Our typology could therefore be extended by production-based material-flow data for cities worldwide. Overall, however, our results provide support for developing differentiated urban mitigation strategies that reflect the variation in the key drivers of urban emissions.

Methods
Correlation Analysis. We performed the correlation analysis for each dataset using a distinct dependent variable: final direct energy consumption per capita for the GEA data, total transportation energy use per capita for the UITP data, and GHG emissions per capita for the WB data. Because dependent variables but also methods of data collection and year of data collection differed, datasets were not harmonized. We calculated the correlation (Pearson correlation coefficient; see SI Appendix) for 10 independent variables: GDP per capita, population density, heating degree days (HDDs; number of days with temperatures <15.5 °C), cooling degree days (CDDs; number of days >23 °C), gasoline price, population size, household size, urbanization level, a center of commerce index (a comparative ranking of 75 of the world’s leading global cities and their instrumental role in driving the economy), and coastal location, a binary variable representing coastal city location. With the exception of coastal location, which did not show any significant correlation with energy use-related variables, results are presented in Fig. 1 and SI Appendix, Table S1. Even though recent studies have normalized GHG emissions of direct energy use with GDP (31, 32), we treat GDP as one of several attributes, which allows a distinction between different classes of cities based on complete bundles of attributes. Each dataset was analyzed independently. Correlation statistics (SI Appendix, Table S1) for each database were calculated individually. Heating degree days (HDDs), cooling degree days (CDDs), and household size were collected independently and were used for all datasets.

Statistical Analysis. We used correlation coefficients as effect sizes to aggregate results across all datasets using the meta analysis random effect DerSimonian-Laird (DSL) approach (33). The random-effects analysis requires first an inverse variance weighting, and then a reverse unweighting by applying a random effects variance component, which is derived from the extent of variability of the underlying studies’ effect sizes. The random effects approach, as opposed to a fixed effect one, permits inference to extend beyond the datasets included in this study (34). It also allows for between-dataset heterogeneity as well as within-dataset variability. Individual studies taken separately tend to consistently underestimate heterogeneity (33).

Regression Analysis. To estimate the relationship between energy consumption and determinants, we used a standard multiple regression approach with log-transformed variables as described in standard applied regression literature (35, 36). The estimated model has the following standard form,

\[ \ln(E_i) = \alpha + \sum_{j=1}^{k} \beta_j X_{ij} + e_i \]

with \( i = (1, \ldots, N) \) representing cities. Here, \( E_i \) is a measure of city energy consumption, and \( X_{ij} \) for \( j = (1, \ldots, k) \), denotes the consumption determinants, with \( k \) the total number of regressors, and \( e_i \) is the classical error term.

The variables included in the final regression were selected from a larger set of possible determinants: i.e., GDP per capita, population density, heating degree days, cooling degree days, gasoline prices, household size, urbanization rate, and commerce center index. The variables included were then reduced, applying the widely used backward elimination statistical procedure (35, 36). The model selection procedure is known as “general to specific approach.” The procedure starts with a large number of variables that are sequentially reduced by removing
the least significant variable, one at the time, if its $P$ value is above a chosen threshold, reestimating the model each time with the remaining variables. The initial selection criterion used was $P > 0.2$ to remove. The procedure stops, when all variables are significant at the 0.2 level. In the UITP dataset, in addition to the other variables reported in Table 1, household size was also significant at the $P < 0.05$ level. Coefficients from a regression model, where the dependent and independent variables of interest are in natural log form and linearly related to each other, can be conveniently interpreted as the average percentage change in the dependent variable corresponding to a percentage change in the independent variable (ref. 37, p. 55). The regression coefficients thus obtained are independent of the units used for measuring variables and are known as elasticities.

**Threshold Regression.** Recursive data partitioning algorithms provide computationally efficient methods to produce the classification that requires processing multiple threshold variables as well as threshold values. In this paper, we used the recursive data-partitioning algorithm developed by Loh and Hansen et al. (38–40), known as GUIDE (generalized, unbiased, interaction detection and estimation) (see also SI Appendix, Section 7), which repeatedly splits the data into increasingly homogeneous groups. The resulting model can be conveniently presented as a binary tree graph. These 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 (see SI Appendix). GUIDE minimizes potential biases in variable selection and interaction detection and allows fitting a linear model at each node.

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 S2
2 Correlation analysis S4
3 Discussion of the emission and energy use elasticities S7
4 Emission/Energy-driven top-down clustering S8
5 Regression with quadratic income term S10
6 Splitting and threshold estimation tree-based method S12
7 Cross-validation analysis to select the size of the tree S14
8 Confidence interval estimation for threshold values S16
9 Peak urban travel S22
10 Calculating the urbanization wedge S22
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 worldwide 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.

\[ \ln(N) = 17.8 - 1.07 \ln(S) \]
2 Correlation analysis

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

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

Where $\text{cov}(X, Y)$ represents the covariance, $\text{var}(X)$ the variance of $X$ and $\text{var}(Y)$ 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.

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

<table>
<thead>
<tr>
<th></th>
<th>Gasoline price vs. population density</th>
<th>Gasoline price vs. linear population density</th>
<th>gasoline price/GDP vs. population density</th>
</tr>
</thead>
<tbody>
<tr>
<td>World Bank</td>
<td>0.52**</td>
<td>0.55**</td>
<td></td>
</tr>
<tr>
<td>GEA</td>
<td></td>
<td>-0.35***</td>
<td></td>
</tr>
<tr>
<td>UITP</td>
<td></td>
<td></td>
<td>0.34**</td>
</tr>
</tbody>
</table>
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 CO₂ 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 than 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.

<table>
<thead>
<tr>
<th></th>
<th>World Bank</th>
<th>GEA</th>
<th>UITP</th>
</tr>
</thead>
<tbody>
<tr>
<td>High emitters</td>
<td>high GDP</td>
<td>med.-high GDP</td>
<td>med.-high GDP</td>
</tr>
<tr>
<td></td>
<td>med.-low density</td>
<td>med.-high HDD</td>
<td>low density</td>
</tr>
<tr>
<td></td>
<td>high HDD</td>
<td></td>
<td>med.-high HDD</td>
</tr>
<tr>
<td></td>
<td>low gasoline price</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium emitters</td>
<td>med.-low GDP</td>
<td>med.-high GDP</td>
<td>med.-high GDP</td>
</tr>
<tr>
<td></td>
<td>med.-low density</td>
<td>med.-high HDD</td>
<td>med. density</td>
</tr>
<tr>
<td>Low emitters</td>
<td>med.-low GDP</td>
<td>low GDP</td>
<td>low GDP</td>
</tr>
<tr>
<td></td>
<td>high density</td>
<td>low HDD</td>
<td>high density</td>
</tr>
<tr>
<td></td>
<td>low HDD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>high gasoline price</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Dataset:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WB</td>
<td>GEA</td>
<td>UTP</td>
</tr>
<tr>
<td>GDP pc</td>
<td>−3.700</td>
<td>1.358*</td>
<td>1.502***</td>
</tr>
<tr>
<td></td>
<td>(5.366)</td>
<td>(0.704)</td>
<td>(0.468)</td>
</tr>
<tr>
<td>Density</td>
<td>−0.378**</td>
<td>−0.070***</td>
<td>−0.551***</td>
</tr>
<tr>
<td></td>
<td>(0.151)</td>
<td>(0.026)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>HDD_{15.5}</td>
<td>0.134**</td>
<td>0.065**</td>
<td>−0.019</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.025)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Gasoline</td>
<td>−0.621</td>
<td>−0.377***</td>
<td>−0.323***</td>
</tr>
<tr>
<td></td>
<td>(0.440)</td>
<td>(0.127)</td>
<td>(0.078)</td>
</tr>
<tr>
<td>(GDP pc)^2</td>
<td>0.195</td>
<td>−0.051</td>
<td>−0.063**</td>
</tr>
<tr>
<td></td>
<td>(0.269)</td>
<td>(0.037)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Constant</td>
<td>21.238</td>
<td>−3.916</td>
<td>3.337</td>
</tr>
<tr>
<td></td>
<td>(26.577)</td>
<td>(3.321)</td>
<td>(2.053)</td>
</tr>
</tbody>
</table>

Observations: 24 (WB), 223 (GEA), 64 (UTP)
R^2: 0.724 (WB), 0.360 (GEA), 0.892 (UTP)
Adjusted R^2: 0.647 (WB), 0.345 (GEA), 0.881 (UTP)

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 UTP 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>
<thead>
<tr>
<th>Node Subset:</th>
<th>( 8 )</th>
<th>( 9 )</th>
<th>( 10 )</th>
<th>( 11 )</th>
<th>( 12 )</th>
<th>( 13 )</th>
<th>( 14 )</th>
<th>( 15 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP pc</td>
<td>0.894**</td>
<td>0.578</td>
<td>0.571***</td>
<td>0.368***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(0.189)</td>
<td>(0.112)</td>
<td>(0.098)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDD</td>
<td>-0.958</td>
<td>-0.240</td>
<td>-0.042</td>
<td>0.145***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.341)</td>
<td>(0.044)</td>
<td>(0.029)</td>
<td>(0.041)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>0.319***</td>
<td>0.466*</td>
<td>0.120*</td>
<td>-0.134***</td>
<td>0.340**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.107)</td>
<td>(0.065)</td>
<td>(0.071)</td>
<td>(0.023)</td>
<td>(0.143)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gasoline</td>
<td>12.084**</td>
<td>4.615***</td>
<td>2.354</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.559)</td>
<td>(1.453)</td>
<td>(0.672)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>3.694***</td>
<td>4.478</td>
<td>2.957***</td>
<td>2.209***</td>
<td>4.087</td>
<td>0.019</td>
<td>0.547</td>
<td>2.400**</td>
</tr>
<tr>
<td></td>
<td>(0.471)</td>
<td>(2.817)</td>
<td>(0.197)</td>
<td>(0.746)</td>
<td>(1.357)</td>
<td>(1.058)</td>
<td>(0.955)</td>
<td>(1.026)</td>
</tr>
</tbody>
</table>

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, Hebei, Nanchang, Shanghai, Shijiazhuang, Taiyuan, Tianjin, Wuhan, Xi'an, Shaanxi, Yinchuan, Zhejiang
  - MAF: Buffalo City, Ekurhuleni, eThekwini, Mmabatho, Nelson Mandela, Potchefstroom
  - REF: Kiev
- **Node 12**
  - ASIA: Dalian, Hohhot
  - MAF: Saldanha, uMhluzane
  - OECD: Bonn, Sydney
- **Node 13**
  - ASIA: Bangkok, Guangzhou, Hangzhou, Hong Kong, Iskandar, Jiaan, 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 Janeiro, Sao Paulo
  - REF: Belgrade, Bratislava, Budapest, Prague, Sofia, Zagreb
- **Node 15**
  - OECD90: Gävle, Helsinki, Jönköping, Karkstad, Linköping, Norrköping, Örebro, Oslo, Stockholm county, Sundsvall, Umeå, Uppsala, Västra Götaland, Växjö
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 \( \{y_i, x_i, q_i\} \) be an observed sample, where \( y_i, q_i \in \mathbb{R} \) and \( x_i = (1, x_{i2}, \ldots, x_{ik})^T \). The threshold variable \( q_i \), which can be an element of \( x_i \), is assumed to have a continuous distribution. The threshold regression model

\[
\begin{align*}
y_i &= \vartheta' x_i + e_i, & q_i \leq \tau \\
y_i &= \theta' x_i + e_i, & q_i > \tau
\end{align*}
\]

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

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

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

\[y_i = \theta^T x_i + \delta^T x_i d_i(\tau) + e_i, \quad (S3)\]

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

\[\hat{\gamma} = \arg \min_{\tau \in G_n} S_n(\gamma),\]

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 \( \hat{\gamma} \) is nonstandard but free of nuisance parameters.

To test the hypothesis \( H_0 : \gamma = \gamma_0 \), a likelihood ratio approach can be employed with test statistic

\[LR_n = n \frac{S_n(\gamma) - S_n(\hat{\gamma})}{S_n(\hat{\gamma})} \quad (S4)\]

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

\[
\hat{T} = \{ \gamma | LR_n \leq c \}. \tag{S5}
\]

Hansen also provides the heteroskedasticity-robust asymptotic confidence set for \( \gamma, \hat{T}^* \), 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, \( \hat{T}^* \) 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 \( \gamma \) 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 $\gamma$ is the value that minimizes the curve, which occurs at about $\hat{\gamma} = $9,700. The 95 % critical value of 7.35 is also plotted (dashed line). The asymptotic 95 % confidence set is $\hat{T}^* = [9,500,13,000]$, 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 $\gamma$ is the value that minimizes the curve, which occurs at about $\hat{\gamma} = 1.2 \$/l. The 95 % critical value of 7.35 is also plotted (dashed line). The asymptotic 95 % confidence set is $\hat{T}^* = [1.07 \$/l, 1.23 \$/l]$, 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 $\gamma$ is the value that minimizes the curve, which occurs at about $\hat{\gamma} = 210 \, \text{pop/km}^2$. The 95% critical value of 7.35 is also plotted (dashed line). The asymptotic 95% confidence set is $\hat{T}^* = [200 \, \text{pop/km}^2, 300 \, \text{pop/km}^2]$, 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 $\gamma$ is the value that minimizes the curve, which occurs at $\hat{\gamma} = 94 \, \text{HDD}$. The 95% critical value of 7.35 is also plotted (dashed line). The asymptotic 95% confidence set is $\hat{T}^* = [70, 670] \, \text{HDD}$, because of the sparcity of the data, is given by the values of $\gamma$ 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 $\gamma$ is the value that minimizes the curve, which occurs at $\hat{\gamma} = 1261$ HDD. The 95% critical value of 7.35 is also plotted (dashed line). The asymptotic 95% confidence set is $\hat{T}^* = [719\text{ HDD}, 2450\text{ HDD}]$, 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 $\gamma$ is the value that minimizes the curve, which occurs at $\hat{\gamma} = 454 \text{ pop/km}^2$. The 95% critical value of 7.35 is also plotted (dashed line). The asymptotic 95% confidence set is $\hat{T}^* = [453 \text{ pop/km}^2, 460 \text{ pop/km}^2]$, 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.
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 $\gamma$ is the value that minimizes the curve, which occurs at $\hat{\gamma} = 2840$ HDD. The 95% critical value of 7.35 is also plotted (dashed line). The asymptotic 95% confidence set is $\hat{T}^* = [2752, 3211]$ HDD, because of the sparcity of the data, is given by the values of $\gamma$ 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 per capita. The least square estimate of $\gamma$ is the value that minimizes the curve, which occurs at about $\hat{\gamma} = $29,300. The 95% critical value of 7.35 is also plotted (dashed line). The asymptotic 95% confidence set is $\hat{T}^* = [22,400, 33,000]$, 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) ±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


