Transport demand in China: Methods for estimation, projection, and policy assessment

by

Paul Natsuo Kishimoto

B.A.Sc., University of Toronto (2008)
S.M., Massachusetts Institute of Technology (2012)

Submitted to the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Engineering Systems

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2018

Author . . . .
Institute for Data, Systems, and Society
August 1, 2018

Certified by . . . .
Valerie J. Karplus
Class of 1943 Assistant Professor of Global Economics and Management
Thesis Supervisor

Certified by . . . .
Jinhua Zhao
Edward H. & Joyce Linde Associate Professor
of Urban Planning and Transportation

Certified by . . . .
John M. Reilly
Co-Director, MIT Joint Program on the Science and Policy of Global Change

Accepted by . . . .
Stephen C. Graves
Abraham J. Siegel Professor of Management Science
Professor of Mechanical Engineering and Engineering Systems
IDSS Graduate Officer

THIS PAGE INTENTIONALLY LEFT BLANK

Transport demand in China: Methods for estimation, projection, and policy assessment

by

Paul Natsuo Kishimoto

Submitted to the Institute for Data, Systems, and Society
on August 1, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Engineering Systems

Abstract

China's rapid economic growth in the twenty-first century has driven, and been driven by, concomitant motorization and growth of passenger and freight mobility, leading to greater energy demand and environmental impacts. In this dissertation I develop methods to characterize the evolution of passenger transport demand in a rapidly-developing country, in order to support projection and policy assessment.

In Essay #1, I study the role that vehicle tailpipe and fuel quality standards ("emissions standards") can play vis-à-vis economy-wide carbon pricing in reducing emissions of pollutants that lead to poor air quality. I extend a global, computable general equilibrium (CGE) model resolving 30 Chinese provinces by separating freight and passenger transport subsectors, road and non-road modes, and household-owned vehicles; and then linking energy demand in these subsectors to a province-level inventory of primary pollutant emissions and future policy targets. While climate policy yields an air quality co-benefit by inducing shifts away from dirtier fuels, this effect is weak within the transport sector. Current emissions standards can drastically reduce transportation emissions, but their overall impact is limited by transport's share in total emissions, which varies across provinces. I conclude that the two categories of measures examined are complementary, and the effectiveness of emissions standards relies on enforcement in removing older, higher-polluting vehicles from the roads.

In Essay #2, I characterize Chinese households' demand for transport by estimating the recently-developed, Exact affine Stone index (EASI) demand system on publicly-available data from non-governmental, social surveys. Flexible, EASI demands are particularly useful in China's rapidly-changing economy and transport system, because they capture ways that income elasticities of demand, and household transport budgets, vary with incomes; with population and road network densities; and with the supply of alternative transport modes. I find transport demand to be highly elastic ($\epsilon_x = 1.46$) at low incomes, and that income-elasticity of demand declines but remains greater than unity as incomes rise, so that the share of transport in households' spending rises monotonically from 1.6 % to 7.5 %; a wider, yet lower range than in some previous estimates. While no strong effects of city-level factors are

identified, these and other non-income effects account for a larger portion of budget share changes than rising incomes.

Finally, in Essay #3, I evaluate the predictive performance of the EASI demand system, by testing the sensitivity of model fit to the data available for estimation, in comparison with the less flexible, but widely used, Almost Ideal demand system (AIDS). In rapidly-evolving countries such as China, survey data without nationwide coverage can be used to characterize transport systems, but the omission of cities and provinces could bias results. To examine this possibility, I estimate demand systems on data subsets and test their predictions against observations for the withheld fraction. I find that simple EASI specifications slightly outperform AIDS under cross-validation; these offer a ready replacement in standalone and CGE applications. However, a trade-off exists between accuracy and the inclusion of policy-relevant covariates when data omit areas with high values of these variables. Also, while province-level fixed-effects control for unobserved heterogeneity across units that may bias parameter estimates, they increase prediction error in out-of-sample applications—revealing that the influence of local conditions on household transport expenditure varies significantly across China's provinces. The results motivate targeted transport data collection that better spans variation on city types and attributes; and the validation technique aids transport modelers in designing and validating demand specifications for projection and assessment.

Thesis Supervisor: Valerie J. Karplus
Title: Class of 1943 Assistant Professor of Global Economics and Management

Jinhua Zhao
Title: Edward H. & Joyce Linde Associate Professor of Urban Planning and Transportation

John M. Reilly
Title: Co-Director, MIT Joint Program on the Science and Policy of Global Change

Acknowledgments

This work was only possible due to the generous support of many people. The people are too numerous to all name, and the debt too extensive to ever fully repay.

I am grateful to Valerie Karplus for her supervision, keen insight at every stage of the work, and unstinting trust when I most needed and least felt I deserved it; an example in advising that I will always seek to remember and emulate. My committee, John Reilly and Jinhua Zhao brought complementary and invaluable perspectives to the work at every stage. In particular, John has showed me how to construct—and deconstruct—economic reasoning, and then communicate it; and Jinhua how to unpack a nigh-endless stream of questions and research ideas from starting points that I mistook for simple and uncomplicated.

Through them, and also through Joe Sussman, John Heywood, and Xiliang Zhang, I found several academic homes and communities at MIT and beyond that have shaped not only the work, but also me. I am grateful for their dedication in creating, seeking funding for, and sustaining these communities. These included the Joint Program on the Science and Policy of Global Change, the Engineering Systems Division, the Tsinghua-MIT China Energy & Climate Project (CECP) and later Karplus Group, the 3E Institute at Tsinghua University, and the JTL Urban Mobility Lab. Each contained peers who motivated me, older colleagues and mentors who volunteered wise advice and encouragement, junior members who asked great questions, and administrative staff who did the tireless work of ensuring we had the time and space in which to puzzle away at research. Funding from the Martin Family through their Society of Fellows for Sustainability, from MISTI, and the CECP and JP sponsors is also acknowledged.

Dava Newman, Noelle Selin, and Ken Oye gave me invaluable opportunities to teach TPP masters' students in the *Leadership Seminar* and *Science, Technology, and Public Policy*; I am thankful for the experience, the funding, the example of their teaching, and their confidence in mine. The brilliant and enthusiastic students helped me realize that instruction can be a source of motivation and energy for the other parts of academic work.

I was lucky to be connected to groups such as the International Transport Energy Modeling (iTEM) community, the ADC70 Transport Energy committee of the Transportation Research Board, the Beijing Energy Network, and the MIT Transportation Club, and to be welcomed warmly by the leaders in those groups.

To follow in the path of my fellow students in the TPP-ESD sequence has been a gift that I could not have conceived of before arriving. It was an immense, once-in-a-lifetime privilege to admire the great and varied successes of my seniors in their early careers; to share the stresses and small victories of the Ph.D. struggle with my contemporaries; and to offer help to those coming up next.

Many friends in Boston, Toronto, Beijing, and around the world—both those who I knew before embarking on this journey, and those I met along the way—have provided a welcome respite from work, while showing undue patience with my alternate busy silences and garrulous unloading of research trivia. Finally, my family—Barb, Steve, Marla, and the entire Kishimoto and Bowen clans—made me the person who was going to set out to do this in the first place, and have seen me through with their love.

Contents

List of Figures

List of Tables

List of Abbreviations

Page numbers are given for the first use of an abbreviation in each chapter.

Notation	Description	
AIDADS	"An implicit direct additive demand system"	92, 162
AIDS	Almost Ideal demand system	4, 83, 161, 191
BRT	Bus rapid transit	19
C-REM	China Regional Energy Model, a global CGE model with provincial detail in China	45, 91, 192
CAFE	Corporate Average Fuel Economy	18
CDE	Constant difference of elasticities	162
CECP	Tsinghua-MIT China Energy and Climate Project	229
CEIC	"China Premium Database" published by CEIC Data	106, 167, 232
CES	Constant elasticity of substitution	91, 162, 196
CGE	Computable general equilibrium	3, 31, 37, 91, 160, 190
CGSS	China General Social Survey	166, 199
CHFS	China Household Finance Survey	199
CHIP	China Household Income Project	97, 160, 197, 215, 230
CLIOS	Complex, large, interconnected, open, and socio-technical (of engineering systems)	17
CRECS	China Residential Energy Consumption Survey	94, 183
DOI	Digital Object Indentifier	229
EASI	Exact affine Stone index demand system	3, 82, 160, 190, 230
ECT	Emissions control technology	36
EPPA	Economic Projection and Policy Analysis model; a global CGE model and component of the IGSM	46, 91, 162, 192
EV	Electric vehicle	29
`FO`	Non-road freight transport sector in C-REM	46
`FR`	Freight road transport sector in C-REM	46
GDP	Gross domestic product	24, 106, 164, 195
GHG	Greenhouse gas	17, 88, 200
GHS	General Household Survey (Nigeria)	166, 199
HDV	Heavy-duty vehicle	36
`HVT`	Household vehicle passenger road transport sector in C-REM	46
IAM	Integrated assessment model	23, 160, 198
ICEV	Internal combustion engine vehicle; fuelled by gasoline or diesel	26

Notation	Description	
IEA	International Energy Agency	23
IGO	International governmental organization; e.g. the IEA	23
IIASA	International Institute for Applied Systems Analysis	200
LDV	Light-duty vehicle	16, 36, 82, 164
LES	Linear expenditure system	162
MESSAGE	Model for Energy Supply Systems and their General Environmental impact	200
NBSC	National Bureau of Statistics of China	47, 104, 231
NDC	Nationally-determined contributions under the UN FCCC Paris Agreement	18, 201
NGO	Non-governmental organization	50
NHTS	U.S. National Household Travel Survey	166
NSS	National Sample Survey (India)	166, 199
PDT	Passenger-distance travelled; in other sources, sometimes PKM or PKT ("vehicle-kilometres [travelled]") or PMT ("vehicle-miles travelled")	21, 142, 193
PEMS	Portable emission-monitoring system	28
PHEV	Plug-in hybrid electric vehicle	162
PM	Particulate matter	17
`PO`	Non-road passenger transport sector in C-REM	46
`PR`	Passenger road transport sector in C-REM	46
rmse	Root mean squared error	171, 197
SAM	Social accounting matrix	47
SEM	Structural equation model	89
SPI	Spatial price index	140
SUV	Sport-utility vehicle	16
TMB	Travel money budget	120
UN FCCC	United Nations Framework Convention on Climate Change	18, 202
VDT	Vehicle-distance travelled; in other sources, sometimes VKM or VKT ("vehicle-kilometres [travelled]") or VMT ("vehicle-miles travelled")	20, 84, 193
VLP	Vehicle license plate	29, 195
WB	World Bank	88

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

The rapid evolution of China's transport system in recent decades has brought greater mobility to more than a billion people. This growth, however, has not been painless. The energy, environmental, and urban impacts of rising transport activity have challenged policymakers to devise responses and manage growth. In turn, researchers have sought to understand the course of growth, its drivers, and options for shaping it; these tasks are complicated by the scope, scale, and speed of change.

To contribute to that effort, this thesis develops improved methods to estimate and project measures of transport system growth—in particular, household demand, energy use, and environmental impacts—and to assess policies aimed at addressing the impacts of growth. In particular, I work at multiple scales to examine how 'micro' differences across households, regions, and sectors relate to macroscale or aggregate measures of demand and its consequences. This chapter motivates the work by briefly introducing (Section 1.1) the impacts, or external costs—from global to local scales—of passenger transport motorization. Section 1.2 describes model-based assessment of these systems, and gives some illustrative examples of models that have been used to study aspects of transport demand growth. Focusing on China in particular, Section 1.3 highlights aspects of its recent history and policy situation, explaining the need for improved methods. With this context established, Section 1.4 outlines the three essays that make up the dissertation.

1.1 Motorization and its impacts

In the United States, the twentieth century opened with the invention of the automobile, and closed with almost universal ownership (800 veh./10^3pers.) of light-duty vehicles (LDVs)—cars, light trucks, and sport-utility vehicles (SUVs) (Oak Ridge National Laboratory 2016). The time between saw radical transformations in the technological, economic, institutional, cultural, and physical context in which households made their decisions to purchase and use private vehicles. These decisions, in turn, entailed changes across society in order to design, produce, fuel, regulate, and normalize private autos, and provide places in which to use them.

As of the early twenty-first century, the LDV markets in wealthy countries are said to have *saturated*: even as incomes rise, auto ownership per capita remains flat; and new vehicles are mainly sold to replace scrapped ones. On the other hand, in low- and middle-income countries, saturation has not occurred. Only a small portion of the population own vehicles, and rising incomes continually bring additional households to the level where auto ownership is not merely desired, but affordable. Where a large portion of the population is near this threshold and economic growth is rapid, motorization has the potential to proceed quickly. This was especially the case in China in the years following the financial crisis of 2007–2008, when vehicle-focused stimulus policies contributed to rates of year-on-year growth in auto sales that peaked at 35 percent. The history of other transport modes has prompted similar comparison. For instance, commercial passenger aviation developed in North America, Europe and the Soviet Union in the 1950s, and traffic grew rapidly with income (Schäfer et al. 2009); Chinese air travel demand only began to grow rapidly in the late 1990's as deregulation coincided with income growth (J. Wang et al. 2016).

Any country or region's demands for travel by these and other modes are emergent attributes of a complex, large, interconnected, open, and socio-technical (CLIOS) passenger transport system (Mostashari and J. M. Sussman 2009; J. Sussman et al. 2005). The passenger transport system is *interconnected* with broader economic-, energy-, and natural (environmental) systems. Viewed at the level of particular countries, it is *open* because technologies, forms of regulation and other elements move and cat across national borders; and because any individual having the desire and means can enter and make use of the system. It is *complex* and *socio-technical* because it implicates not only market transactions and physical vehicles, but people with preferences and culture, institutions of government, firms with certain business practices, and the physical (especially urban) spaces in which travel takes place. This engineering systems perspective makes clear that the process of household motorization, and more broadly transport system growth, will differ from time to time and place to place.[1]

1.1.1 Impacts of motorization and policy responses

The impacts of transportation demand range from the very local to the global. A short overview of two areas related to the work in this thesis will illustrate the complexity of impact channels and policy responses. Note that the impacts often fit the rubric of an *externality*: a cost imposed on some other party, not faced or paid by the person whose transport activity creates that cost.

Energy use and emissions. Vehicles burning gasoline or diesel fuel emit both greenhouse gases such as carbon dioxide (CO_2), and other chemicals that are precursors of air pollutants such as particulate matter (PM, e.g. $PM_{2.5}$ and PM_{10}) ($PM_{2.5}$

[1] Changes do not always appear first in higher income countries: for instance, "dockless" or "free-floating" bike sharing technology, in which individuals use a smartphone app to locate, unlock and pay for trips on a fleet of shared bicycles, was pioneered in China in 2014, underwent rapid expansion in 2017, and is only beginning to spread to the United States and Europe, where older, docked technology had already been widespread (Fishman 2016; Parkes et al. 2013).

and PM_{10}) and ozone (O_3). The former contribute to global climate change, while the latter affect the health of individuals breathing polluted air, both nearby and in places to which pollutants are carried by the atmosphere.

Policy responses to these impacts occur at every level. For instance, countries are parties to the United Nations Framework Convention on Climate Change (UN FCCC) and submitted nationally-determined contributionss (NDCs) towards the 2015 Paris Agreement, describing actions that they would take towards the goal of avoiding the worst impacts of climate change. For many, these actions included policies such as fuel economy standards to reduce the fuel burned by road vehicles; the supply or promotion of low-energy, low-emissions modes like mass transit and active transport; and electrification of vehicles. These responses change but do not entirely eliminate the impacts targeted: for instance, vehicles powered by electricity currently mostly rely on coal- or gas-fired generators, so while they are clean at point of use, they still contribute indirectly to global and local pollution. The economic costs of responses also affect their adoption and impact. Fuel flexibility—the option to swap carbon-intensive for low-carbon energy sources—has raised the prospect of deep decarbonization of electric power generation and industry, yet transport sector climate policy action has lagged, in part because fuel-switching for vehicles was expected to be expensive (Creutzig et al. 2015; Gota et al. 2016; Knittel 2012; Pauw et al. 2018).

The impact of air pollution was not initially linked to early motorization and transport; in the United States, recognition of the problem did not occur until the 1950s (Haagen-Smit 1952). Responses included the first tailpipe standards on pollutant emissions (1966) and formation of an Air Resources Board (1967) in the state of California; these institutional factors mean that transport environmental policy continues to be set at both national and sub-national levels. Separately, concern over dependence on imported energy and the costs of securing supply led to the passage of national Corporate Average Fuel Economy (CAFE) standards in 1975. These standards have been progressively tightened. Other cities, countries and sub-national jurisdictions have also responded with to air quality problems associated with vehicle emissions and other sources with a complex set of interacting policies (L. T. Molina and M. J. Molina 2002).

Urban impacts. Drivers' vehicles occupy space on roads; when the design capacity of a road is exceeded, a time cost is imposed on other road users as their travel is slowed. The unpriced externality leads to inefficient allocation of road space—in short, traffic congestion. Along with the impacts of local vehicle emissions, vehicles create noise; may injure pedestrians or cyclists; and must be parked when not in use (Gärling and Steg 2007).[2]

These impacts prompt a very wide variety of responses from local governments. To highlight this variety with only a few selected examples, transit-oriented land-use planning aim to reduce the need for vehicle trips by siting or encouraging residences, workplaces, and amenities where they can be connected on foot, bicycle, or by public transit. Numerous forms of direct pricing are in use: of parking spots; and of road use via automated toll systems with dynamic rates. In contrast, desirable transport activity, such as the use of public transport or bicycle share systems, may be subsidized with discounts or free travel for some or all riders. Cities invest in the construction of mass transit—subways, regional rail, trams, and bus rapid transit (BRT)—and in expansion of bus fleets; but also in the building and widening of roads and other infrastructure for LDVs. Police and other officials are hired to enforce laws and regulations.

[2]Research in parking policy highlights that vehicle parking is a land use that is efficient only when its cost is directly comparable to alternate uses, e.g. housing, or commerce.

1.2 Model-based assessment of transport system evolution and policy

In order for governments to inform decisions about such policies, and for firms and other actors to engage with policy processes, they rely in part on model-based knowledge about transport activity growth and its response to context.[3] We can think of a model as encoding a functional relationship, $f(\cdot)$, between certain groups of quantities:

$$(Y, Z) = f(A(t, X), X)$$

These include a vector, A, of measures of transport activity varying with time, t: for instance, the vehicle-distance travelled (VDT) in certain types of vehicles; the number of trips by other modes; amounts of money spent on certain types of travel; quantities of freight goods moved, etc. Policymakers target a set of outcomes, Y, such as levels of traffic congestion on certain roads, or total GHG emissions from vehicles; and stakeholders may be concerned with a second vector, Z, of additional effects of transport activity. The policies mentioned above all have design parameters, X, that influence activity directly and/or mediate its impacts. For instance, the parameters for vehicle fuel economy standards include the amounts of fuel allowable per unit distance driven; the types and sizes of vehicles to which these amounts apply; dates at which the standards will tighten; and the size of penalties to encourage compliance. For road pricing or tolling, parameters include the prices; their basis (trip or unit distance); their variation over time; and the roads to be covered. Stakeholders seek to know:

— Without a certain policy or decision, what would be the counter-factual activity, A, or impacts, Y, be?

— Given a new or revised policy, X, what will be the resulting demand, A, and

[3] by *context*, I denote the other transport and non-transport systems to which a particular transport system is (in the CLIOS sense) *open* and/or *interconnected*; and the *socio-technical*, or institu-

impacts, Y?

— What side-effects will there be on other impacts of secondary interest, Z?

The outcomes of interest and policy levers differ across stakeholders. For instance:

— Auto manufacturers and their suppliers size investments in facilities, workforces, and their supply chain in order to be profitable in a future vehicle market of uncertain size and composition across segments (small cars, SUVs, etc.).

— An international development bank investing in an intermodal transport facility—e.g. an airport with rail and road access and parking—will be concerned that it is sited, sized, designed, and operated appropriately to meet future demand.

— Countries considering collective progress on achieving the mitigation goals of the Paris Agreement seek to understand how much one another's transport GHG emissions might grow in the future, and the level of mitigation effort represented by transport-sector actions in each country's NDCs. They use this information to pressure one another to take more action, or to choose their own level of effort.

1.2.1 Models of transport systems and transport concepts

As researchers study transport systems using models, they focus on measures or indicators of the background concepts (transport system attributes) of interest (Adcock and Collier 2001). For instance, demand can be measured by numbers of trips; by passenger-distance travelled (PDT) or VDT, or by expenditure on travel. Measures are chosen to be salient to particular research and policy questions, and with attention to the data available for modeling. The choice of measures in turn affects the type of knowledge produced. Because resources are limited, these choices also affect both the scope and scale of models built and used. Some examples (Table 1.1 on

tional framework that governs it.

Table 1.1: Units of analysis, scope and levels of resolution in models including transport and transport energy.

Scope @ Resolution			
Space	Time	Data source(s)	Examples
Global total	Century @ 1–10 years	Summed national statistics	DICE (Nordhaus and Boyer 2000)
Global @ country or country group	Decades @ 1–5 years	Aggregated from national statistical bodies, via IGOs or input-output databases	EPPA (Chen et al. 2015), GCAM, MESSAGE (McCollum et al. 2017), IEA MoMo
Global or national @ state/sub-national unit	Decades @ 1–5 years	National statistical bodies	C-REM (D. Zhang, Rausch, et al. 2013), USREP (Rausch et al. 2011)
Subset of all cities in one or more countries	<10 years @ 6–12 months	City-level statistics or databases thereof; directly, via national statistical bodies or NGOs	ITF-OECD urban mobility model, S. Wang and J. Zhao (2018)
One or a few cities @ households or individuals (sampled)	1–7 days @ 10–60 minutes	Travel surveys, passive data	Agent-based (Waraich et al. 2015); demand econometrics; machine learning

page 22) will illustrate the diversity of models applied to transport systems.

Agent-based models. Researchers of transport activity in cities use individual people as their units of analysis, and simulate agents' location along particular road segments to the minute or second. Such models are suitable for investigating how changes in specific transport network links, such as widened roads or new public transit stops, affect travel patterns; or how diurnal patterns of activity give rise to congestion during rush hour periods. For such modelers, the costs of data collection, processing, and computation also limit scope, both in the period of time (days and weeks instead of years and decades) and space (single cities instead of entire countries) that can be simulated (Waraich et al. 2015).

Integrated assessment models. At the far end of the scale, global integrated assessment models (IAMs) are used to assess climate policy questions—including about the effects of long-run economic growth and energy transitions on transport activity, the resulting contribution of transport to climate change, and its role in mitigation vis-à-vis other sectors. Consequently, the models are designed for broad scope: coverage of the entire world; of the remainder of the economy besides transportation; and a time period of decades, because the total emissions rate of greenhouse gases is an essential measure when studying climate change, and its impact extends into the future.[4]

Current models use countries or country-groups and sectors as their units of analysis, and annual- or multi-year time resolution. Data on total consumption of fossil fuels is related quantities is collected and published by national governments, for regulatory purposes including securing supply and limiting consumption and emissions. These national data can in turn be systematized in a consistent manner by international governmental organizations (IGOs) like the International Energy Agency

[4]Since the primary GHG, CO_2, is a well-mixed and long-lived gas, a tonne of CO_2 eventually

(IEA), and linked to measures of GHG emissions. It is therefore low-cost to obtain and use in models.[5] Consequently, existing IAMs are limited in their ability to reflect and capture heterogeneity at levels smaller than the modeled units, including economic subsectors (such as individual transport modes, technologies or vehicles), and human populations (such as specific cities, households, or individuals within model regions). As transport systems grow and evolve, modelers carefully adjust parameters and data at the country level to reflect changes that occur at finer resolution—for instance, policy impacts, new travel options or types of vehicles, and changing consumer preferences. These tuning and updating procedures allow large-scale models to incorporate new information, and thus produce more reliable projections.

Consumer adoption and growth curves. Other modelers seek to find and apply relationships at different levels of observation and analysis. Dargay et al. (2007) give an example of the widely used approach of fitting logistic (Gompertz) curves (Equation (1.1) on the next page) to data at the country-year level of resolution (Figure 1-1 on the facing page). Drawn from the literature on consumer adoption of new technologies such as refrigerators, cameras, and mobile phones (Bass 1969; Bonus 1973; Golder and Tellis 1998), these are a family of 'S'-shaped functions encoding the conceptual relationship, mentioned on page 16, between growing wealth or gross domestic product (GDP), and motorization. Many stakeholders focus on and use such functions because they directly encode and output the stock and sales of LDVs, and thus help to inform decisions related to the manufacture, marketing, trade, and regulation of vehicles. Huo and M. Wang (2012) and T. Wu et al. (2014) and others

contributes the same amount to climate change whether emitted in emitted in Argentina or Zambia. Emissions rates differ in important ways by country, by province, by city, or by individual source, but an accurate global total is indispensable for climate modeling.

[5] If a researcher sought to use firms, instead of countries, as units of measurement in an IAM, they would find that individual firms measure energy use and emissions in idiosyncratic ways (if at all). The resulting data would be costly to harmonize. It would be incomplete, missing portions of the global total, so that the basic sum would need careful adjustment, entailing further effort. The researcher would also find a smaller literature to rely on in making these adjustments, as the cost would have dissuaded many others from making the attempt.

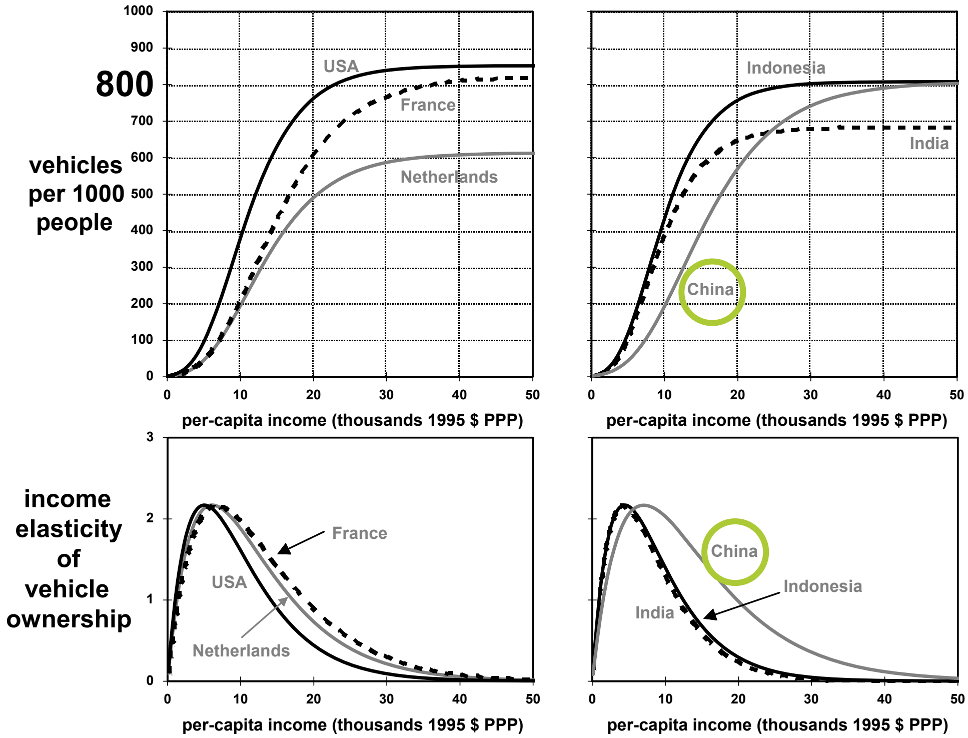


Figure 1-1: Figure 8 from Dargay et al. (2007). The authors adjust the saturation level of per capita ownership in China (upper right panel) below the U.S. level (upper left panel) using national data on density and urbanization. For discussion of income elasticities of demand (bottom panels), see Section 3.2.1 on page 85.

have applied these relationships to projecting motorization in China.

$$\text{Vehicles / capita} = \gamma \exp\left(\alpha \exp\left(\beta \times \text{GDP / capita}\right)\right) \tag{1.1}$$

As with IAMs, careful research that estimates and uses such relationships will include adjustments for context. For instance, in Dargay et al. (2007), the parameter for vehicle ownership per capita at saturation, γ, is picked to be the highest level observed in the data—the United States. The authors set the effective γ in each country by decreasing it using factors for population density and share of the population that is urbanized, with country-specific data and one global coefficient for each variable. Thus Figure 1-1 shows saturation in China at roughly 800 veh./10^3pers., instead of the higher 850 veh./10^3pers. in the U.S.

1.2.2 Challenges in transport system modeling

Along with the need to adjust for or extrapolate changes outside model scope (for models such as agent-based models), or below model resolution (for IAMs or growth-curve models), two further obstacles create problems in using models to project transport demand and assess transport policies. First, different models and frameworks often rely on alternate functional relationships and underlying logic. Lack of consensus on which are most appropriate across contexts is an instance of structural uncertainty, and results in divergent projections. For instance, Figure 1-2 on the facing page from Yeh et al. (2016) shows that four global models that include transportation and energy predict different motorization paths for China from a 2005 base year. The authors link low projected vehicle ownership from one model (GCAM) to a low projection of overall mobility (PDT per capita) as incomes rise; other models use different structures that do not explicitly forecast PDT (MoMo, Roadmap) or a similar structure with distinct parameterization and calibration (MESSAGE).

Second, it may be that the core modeling methods do not permit essential adjustments for a particular context. For instance, regarding logistic curves of GDP and vehicle ownership per-capita, Hsieh et al. (2018) show that early-stage motorization data at the national level in China do not contain much information about the eventual saturation level of ownership (Figure 1-3 on the next page), a key parameter in that relationship; and so precise projections are impossible without choosing a Chinese saturation level based on assumed similarity to other countries. Adjustments to relationships modeled at the country level may also founder when system change occurs heterogeneously across units at finer resolutions. For instance, as the availability of ride-hailing services and, eventually, autonomous vehicle technologies spreads from city to city, the relationship of aggregate vehicle ownership to other quantities will shift. A similar problem may arise if individual cities make plans, as Paris did in 2017, to ban certain types of vehicles such as internal combustion engine

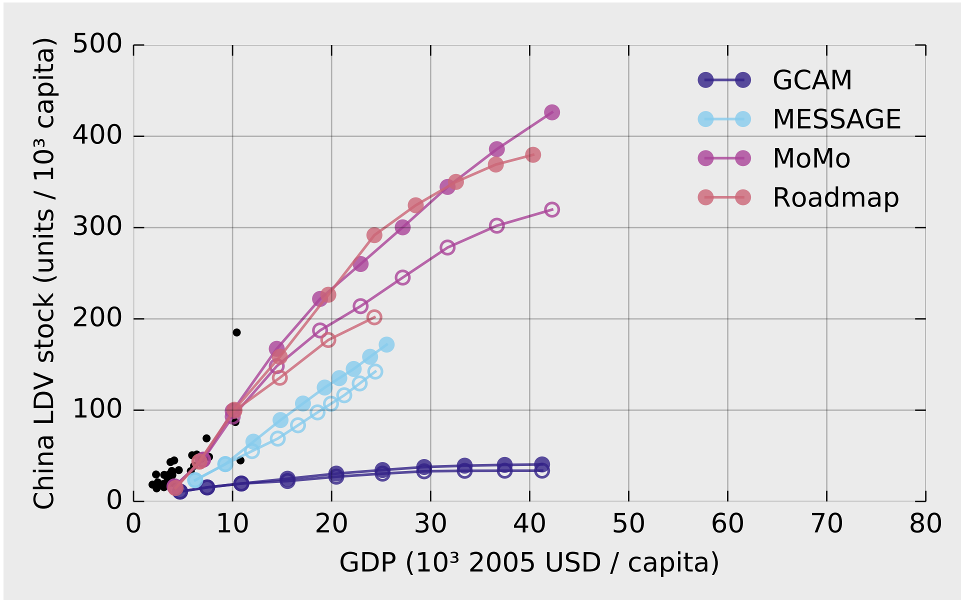


Figure 1-2: Projections of China national GDP and vehicle ownership, both per capita, from four global models (Yeh et al. 2016, re-plotted from supplemental data); five-year increments with reference scenarios as filled marks. Black marks: data for individual Chinese provinces as of 2010 (National Bureau of Statistics of China 2012, black dots).

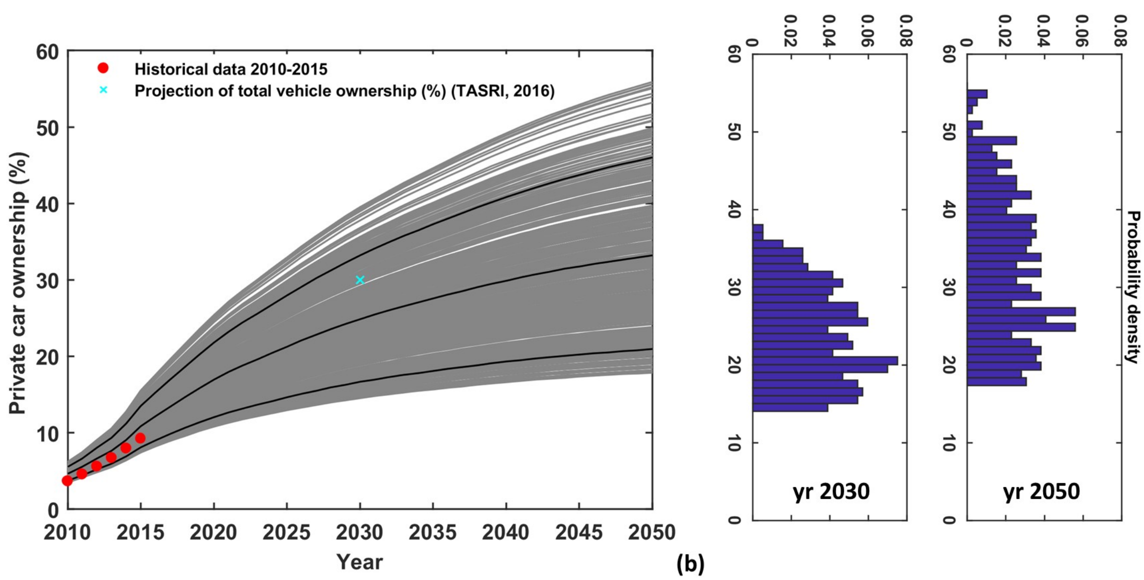


Figure 1-3: Figure 4b from Hsieh et al. (2018). Grey lines: 400 samples from a Markov Chain Monte Carlo simulator of the Gompertz parameter distribution, with historical data as priors. This quantification of uncertainty gives a distribution of equally likely fits to historical data, showing that the saturation level of per capita ownership is not narrowly constrained by available data.

vehicles (ICEVs). Without explicit methods to incorporate such changes, which affect individual sub-units differently, model-based assessments risk producing inaccurate conclusions.

1.3 Transport system growth and policies in China

China's size, diversity, and rapid transport system evolution pose exactly these challenges to model-based assessment. Sperling and Gordon (2008), M. Wang et al. (2006), and Z. Zhao et al. (2013) (among many others) and Figure 1-2 reflect the startlingly high growth rates in vehicle sales attained circa 2010, after which China became the world's largest single vehicle market. Growth in other modes has been as rapid. The relationship with income growth is complex: the central government has built or funded over 2.5×10^4 km of high speed rail lines in the last 15 years, an even longer set of expressways, as well as airports and other infrastructure (Hou and S.-M. Li 2011; J. Wang et al. 2016). Beyond supplying transport options, the large scale of investment was a factor in the very GDP growth that raised incomes to the point that some households were able to motorize.

At the same time, China's policy environment is distinct in its configuration due to its form of government, and highly heterogeneous. Motorization in China was minimal until recent decades, and thus regulation of emissions from LDVs is a relatively recent phenomenon; responses to the energy, climate change, and local air pollution impacts of LDVs did not emerge until these problems became severe and unavoidable (Ho and Nielsen 2007; Nielsen and Ho 2013). Yet in recent years, China's nationwide standards on fuel economy, tailpipe emissions, and fuel quality have been tightened rapidly, and are converging with the most stringent examples from other countries (Z. Yang and

[6] In some areas, such as the use of portable emission-monitoring system (PEMS) to validate laboratory test data against the real-world performance of vehicles, China's standards may soon move ahead of those in other jurisdictions.

Bandivadekar 2017).[6] On the other hand, China's current per capita auto ownership is only at the level of the USA circa the 1930s (Oak Ridge National Laboratory 2016; M. Wang et al. 2006); a time when vehicle regulation was minimal and did not impact ownership growth. In addition to these standards, Chinese governments at the central, provincial and municipal level have implemented policies targeting 'electric-first' motorization;[7]; constraints on the overall number of vehicles; and restrictions on vehicle use (both discussed below).

The governments of China's cities each face individual patterns of transport system evolution, and have priorities distinct from the central and provincial governments. Addressing local externalities may be more important, to cities, than national or regional priorities such as, e.g., the support of a strong, export-capable auto manufacturing industry. As well as avoiding the negative impacts of motorization mentioned, cities seek to create places that attract residents and businesses, leading each to adopt a unique combination of transport policy instruments including those listed in Section 1.1.1. Knowledge about policy options and best practices is transmitted between cities through a complex network of officials and experts (J. Zhao and Z. Wang 2014).

One example of a city-level policy response that has seen widespread adoption in China, yet not elsewhere, is the vehicle license plate (VLP) quota, whereby a city limits the number of license plates provided monthly to prospective vehicle owners. The available plates are allocated by lottery, auction, or some hybrid of these methods. First adopted in China by Shanghai in 1994, following the example of Singapore, since 2010 the policies have spread to affect more than 100 million urban residents in seven cities, with at least eight more considering such restrictions (Table 1.2 on the following page). Municipal governments in China are not elected and so, despite opposition from would-be vehicle owners, these policies are maintained and strongly enforced, forming a constraint on households' transport decisions. Some households

[7] in which new buyers are encouraged to purchase an electric vehicle (EV) of some sort, instead of ICEV that is later replaced or supplemented with an EV.

Table 1.2: Summary of policy details across Chinese cities implementing (panel B) or floated as considering (panel C) ownership restrictions. Population in millions for metropolitan or urban area.

City	**Dates** Hinted	Announced	Begun	**Type**	**Pop.**
		— **A** —			
Singapore	—	—	*1990-05-01*	*Auction*	
		— **B** —			
Shanghai	—	—	1994-01-01	Auction	24.3
(Feng and Q. Li 2013)	100–120k/year.				
Beijing	2010-12-13	2010-12-23	2010-12-26	Lottery	21.5
(J. Yang et al. 2014)	Initially 240k/year; 150k/year from 2014; 6m total cap by 2017. 88% private.				
Guiyang, GZ	—	—	2011-07-11	Lottery	4.3
(China Daily 2014; Mu 2011)	24k/year + unlimited suburban.				
Guangzhou, GD	—	—	2012-07-01	Mixed	13.1
(H. Zeng 2012)	120k/year. 40% auction, 60% lottery (of which 10% NEV).				
Tianjin	2013-08	2013-12-15	2013-12-16	Mixed	15.2
(Y. Cheng 2013; Qing 2013)	100k/year. 40% auction, 60% lottery (of which 10% NEV). 88% private, 12% commercial, 0% government.				
Hangzhou, ZJ	2013-07	2014-03-25	2014-03-26	Mixed	8.8
(Xinhua News Agency 2014a)	80k/year. 20% auction, 80% lottery.				
Shenzhen, GD	2013-07	2014-12-29	2014-12-29	Mixed	10.6
(Xinhua News Agency 2014b)	100k/year (20k NEVs). 50% auction, 50% lottery.				
Shijiazhuang, HE	2013-07	2013-06	2015-??-??	—	12.8
(F. Li 2014)	2.1m total @ 2015-12-31; 2.5m total @ 2017-12-31 ~ 200k/year. $\leq$ 2 cars/household.				
		— **C** —			
Chongqing (F. Li 2014)	2013-07	—	—	—	17.8
Qingdao, SD	" "	—	—	—	8.7
Chengdu, SC	" "	—	—	—	14.0
Wuhan, HB	" "	—	—	—	10.2
Changchun, JL (Xue 2015)	2015-04	—	—	—	7.6
Harbin, HE	" "	—	—	—	6.7
Baoding, HE	" "	—	—	—	2.2
Lanzhou, GS	" "	—	—	—	3.6

that reach a level of income at which they might wish to purchase and use a vehicle are prevented or delayed by such quotas. This is one example of a response that alters the connection between economic growth and vehicle purchase decisions in a way that varies from city to city—thereby challenging both the use of observed data to describe that relationship, and of models to project future purchases.

Overall, the rapid changes in China's transport systems have prompted a variety of efforts—model-based and otherwise—to describe the drivers of past growth, project its future course, and help support policy responses. In each of the essays that follow, I survey relevant parts of this literature.

1.4 Outline of the dissertation

Researchers can choose to model China's transport demand, energy use and environmental impacts using relationships gleaned and data collected at different levels of resolution. The foregoing considerations raise the central question: what insights and analytical capacity are gained, and at what cost in difficulty, by moving from national aggregates to greater detail? Conversely, in order to study the impacts of important contextual aspects—policy instruments, or geographical heterogeneity in transport systems—what level of resolution is necessary in models and data?

This thesis contributes methodological responses to this broad challenge, organized into three essays in two areas of work. First, in Chapter 2, I develop methods for increasing geographical and sectoral detail in aggregate computable general equilibrium (CGE) models of China's economy, to both increase realism by carrying forward variations at these levels of resolution, and to bring units of analysis to levels which better match the targeting of existing policy.

Second, in Chapters 3 and 4 together, I develop a novel application of recent econometric methods, in order to derive empirical facts about demand from social science survey data. I show how these data—collected without any particular focus on transport—can be used within flexible demand systems to capture how the transport share in Chinese households' expenditure varies across the income spectrum, and in relation to measures of transport system and urban context at the city level. Finally, I cross-validate the new demand systems across geographical subsets of data, and against more widely-used formulations, to understand how what advantage they confer, where they are sensitive to partial coverage in the data used for estimation.

Throughout the work, I maintain a perspective of transport in the context of broader economic activity. In Chapter 2, the CGE framework incorporates the entire global economy, intermediate demand and output of other economic sectors. In Chapters 3 and 4, though I focus on household consumption and specifically the category of transport expenditures, the systems of demand simultaneously account for other uses to which households may put their money, letting these constrain transport budgets.

Emerging from these innovations, I provide findings, in the first part, on the course of future transport demand growth absent policy, and analyse the interaction of climate and transport-sector policy instruments. The second part, gives new evidence of the transport spending behaviour of Chinese households, and its relationship to factors such as population density and the supply of transit infrastructure, in the period 1995–2007. I also make contributions to modeling practice, by describing how these new demand systems may be applied in projection, simulation, and policy analysis; and to future data collection efforts, by identifying dimensions of variation that should be spanned in order to capture the range of household behaviour.

Figures and tables are included inline where possible. Supplemental figures and tables are placed in chapter appendices or, where very long, in Appendix A. Each chapter contains its own references; the bibliography on page 202 serves for both this chapter and the conclusory Chapter 5. Towards the goal of reproducibility,

Appendix B details the software, data and other materials necessary to recreate or build on the research.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2

The impact of coordinated policies on air pollution emissions from road transportation in China

Abstract

Improving air quality across mainland China is an urgent policy challenge. While much of the problem is linked to China's broader reliance on coal and other fossil fuels across the energy system, road transportation is an important and growing source of air pollution. Here I develop analytical capabilities for studying province-level impacts on total air pollution of implementing vehicle emissions standards, and for comparison to broader, economy-wide climate policy, inside a computable general equilibrium (CGE) model.

I find that full and immediate implementation of existing vehicle emissions standards at China 3/III level or tighter will significantly reduce the contribution of transportation to degraded air quality by 2030. I further show that transportation emissions standards function as an important complement to an economy-wide price on CO_2, which delivers significant co-benefits for air pollution reduction that are concentrated primarily in non-transportation sectors. Going forward, vehicle emissions standards and an economy-wide carbon price form a highly effective, coordinated policy package that supports China's air quality and climate change mitigation goals. The methods developed also support atmospheric modeling and integrated assessment of transport policies.

2.1 Introduction

Air quality is exacting a rising toll on human health and quality of life in China. A broad variety of policy measures have been announced, and some enacted—these include increasing monitoring and reporting to understand the scope and spatial/temporal nature of the problem; setting technology standards; assessing fines and pollution charges; and directly influencing the economic activities which produce pollutants as a byproduct.

Transportation is the target of an important subset of these policies. Fossil fuels (gasoline and diesel) burned in road vehicles (cars, trucks, buses, taxis, etc.) result in direct emissions of pollutants, including those listed in Table 2.1 on the next page. These direct emissions mix with emissions from other large combustion sources—especially electric power plants and industrial facilities—and affect ambient concentrations of pollutants such as fine particulate matter ($PM_{2.5}$) and ozone (O_3), which in turn impact human health.

Transportation sector policies—summarized in Section 2.2.2—include standards regulating the allowable tailpipe emissions of specific pollutants from new private passenger vehicles (light-duty vehicles (LDVs)), and heavy-duty vehicles (HDVs) including light-, medium- and heavy trucks for freight transport, and buses for passenger transport. These standards may be set to promote installation of specific technology, such as diesel particulate filters (DPF), for compliance. Impurities in gasoline and diesel fuel are also regulated, to ensure that these emissions control technologys (ECTs) can function. Collectively, the combination of road vehicle tailpipe and fuel quality standards are referred to as 'emissions standards' (ES).

At the same time, China's broader climate and energy policy agenda has important implications for air quality. The U.S.-China Joint Announcement on Climate Change in November 2014, and China's subsequent pledged contribution to global climate mitigation efforts targets a reversal of rising CO_2 emissions at latest by 2030.

Table 2.1: Primary pollutant species in this analysis, and other species included in the Regional Emissions in ASia (REAS) database, version 2.1 (Kurokawa et al. 2013), used as the basis for emissions projections in this study. 'VOCs' are volatile organic compounds.

Name	**Chemical formula**
This analysis:	
Black carbon	BC
Carbon monoxide	CO
Nitrogen oxides	NO_X
Organic carbon	OC
Sulfur dioxide	SO_2
Also in REAS 2.1:	
Methane	CH_4
Carbon dioxide	CO_2
Nitrous oxide	N_2O
Ammonia	NH_3
Non-methane VOCs	NMV
Particulate matter $\leq$10 µm	PM_{10}
Particulate matter $\leq$2.5 µm	$PM_{2.5}$

Achieving this goal will require economy-wide policies, such as a CO_2 price, which is currently being piloted in some regions and is expected nationwide within the next five years. Climate policy and vehicle emissions policies will both act on the energy and transportation system, with important implications for future air pollution emissions and air quality outcomes.

To better understand how these policies will act together to affect future air pollution in China, I augment an energy-economic computable general equilibrium (CGE) model to support simulation of the combination of road transportation emissions standards and an economy-wide CO_2 price. Section 2.2 surveys past research and existing policies, and describes scenarios combining both types of these policies at varying levels of stringency. Section 2.3 explains how road freight, road passenger, and household vehicle transport are disaggregated in the CGE model using province-level data; in Section 2.4, modeled energy use is linked to a provincially-resolved inventory of historical emissions, as well as standards planned for future road vehicles.

In Section 2.5, I find that ES are projected to be highly effective in reducing the total quantity of emissions from road vehicles, despite rapid growth in transportation activity to 2030—especially when these policies are deployed and, importantly, enforced in an accelerated manner nationwide. This deployment will be important as the demand for passenger and freight vehicle travels grows, and associated emissions increase from a small share of the total today to a much more substantial share. I further find that an emissions standard is complementary to economy-wide climate policy that reduces CO_2 in sectors where the marginal costs of its abatement are lower, and delivers substantial co-benefits in the form of air pollution reduction.

Since the least cost opportunities to reduce CO_2 are mainly concentrated outside of the transportation sector, an emissions standard that directly targets pollution in the transportation sector delivers a significant additional contribution to air pollution reductions. Thus a CO_2 price plus vehicle emissions standards function as effective and complementary coordinated strategies for addressing air pollution and climate change in China.

Finally, Section 2.6 discusses how the methods described here can support integrated assessment of the health and economic impacts of policies on the air pollutant emissions of transportation; presents policy recommendations; and describes how the work can be extended.

2.2 Modeling transport emissions and policy

2.2.1 Literature

Integrated assessment of policy co-benefits. The phenomenon of air pollution co-benefits of climate and energy policy has been long recognized and studied, including in Europe (Harmelen et al. 2002; Nam, N. E. Selin, et al. 2010; Rive 2010) and more recently in China (Aunan et al. 2004; K. He, Lei, et al. 2010; Nam, Waugh, et al. 2013). In particular, Nam, Waugh, et al. (2014) applied economy-wide, general-equilibrium models to compare the potential co-benefits in the U.S. and in China, in light of contrasting energy systems, and the stringency of existing control measures.

Air pollution from the transport sector. For assessment of pollution and health impacts within the transport sector, S. Yang and L.-Y. He (2016) modeled individual Chinese provinces as independent economies, using regression models and "pollution elasticities" to estimate health effects under future fuel price scenarios. L.-Y. He and Qiu (2016) took a similar approach for the country in aggregate, but studied instead the effect of mode shifts. X. Wu et al. (2016) used provincial-level modeling to assess emissions control policies 1998–2013, concluding that continued growth from heavy-duty (especially diesel) vehicles and enforcement of type approvals were areas of key concern. On the side of fuel quality, Yue et al. (2015) sampled fuels at about 60 sites in 2010–2011, discovering significant variation and exceedances, and suggested policy adjustments to promote compliance.

Guo et al. (2016) compared the projected effects of four transport-sector policies applied to the Beijing-Tianjin-Hebei (or Jing-Jin-Ji, JJJ) region, including accelerating the adoption of ES; and S. Zhang et al. (2016) similarly designed strategies for cities in the Yangtze River Delta. Lang et al. (2012) studied JJJ retroactively for the period 1999–2010, noting that increases in freight traffic were related to increases in transport NO_x and PM_{10} emissions even as other species decreased. H. Wang, Fu, et al. (2010) developed 1995–2005 inventories of vehicle emissions for the large cities of Beijing, Shanghai and Guangzhou. Lu et al. (2017) focused on school trips in particular, noting their contribution to congestion and pollution, and Deng (2006) measured the monetary costs of vehicle-related pollution in Beijing by two econometric methods.

Emissions from vehicles and uncertainty in inventories. Emissions from road vehicles have been studied by a variety of methods. Following in-use vehicles on actual roads with specialized instruments, Huang et al. (2016) measure the emissions from bi-fuel vehicles, while Zheng et al. (2015) present an instrument and drive-cycle methods focused on black carbon (BC).

For inventories of total, rather than specifically transport, emissions, studies such as Hu et al. (2015), use such direct measurements and bottom-up accounting to drive emissions inputs to atmospheric simulation models, aiming to reproduce changes in observed secondary air pollutant (i.e., $PM_{2.5}$ and O_3) concentrations. S. Cheng et al. (2013) developed a hybrid approach incorporating ground monitoring data, focusing on Beijing only. Miyazaki and Eskes (2013) used satellite measurements and assimilation techniques to constrain the estimates of Kurokawa et al. (2013) (the REAS inventory used in the present study).

R. Wu et al. (2016) developed a bottom-up inventory for VOCs only at the province level, including the contribution of road vehicles. Hong et al. (2016) focus on the contribution of uncertainty in energy statistics to bottom-up methods, finding high ratios of maximum discrepancies to mean values—for instance, the total 2012 inventory of SO_2 emissions may vary up to 30%, and NO_x by 16.4%, due to energy use uncertainty alone.

Finally, Xia et al. (2016) combined satellite data with bottom-up estimates to assess the effects of industrial- and power-sector policies during the 11th (2006–2010) and 12th Five-Year Plan (2011–2015) periods, noting the growing contribution over this period of NO_x from transportation.

In terms of assessing past and future changes in transport activity—due to both growth, and policy—prior studies have focused on different geographies or aggregations, transport modes, policies, and modeling methods, yet have generally said little about the relationship with policies not focused on the transport sector. Conversely, examination of the co-benefits of climate policy has tended to focus on economy-wide impacts, or comparison with specific measures in the power- and industrial sectors, rather than the transport sector. The present work bridges this gap by providing methods for assessing both the climate policy co-benefit of air pollution reduction, and road transport emissions reductions due to ES, in a consistent, economy-wide framework.

In doing so, I note that researchers continue to work to resolve the uncertainties in the history and current state of vehicle tailpipe emissions. These are relevant to our method for deriving transport-subsector-specific emissions factors from a database (Kurokawa et al. 2013) that also covers the non-transport sectors where climate policy co-benefits also arise; a matter taken up further in Section 2.4.1.

2.2.2 Existing policies and policy scenarios

To investigate transport-sector emissions standards, the magnitude and distribution of their impact can be compared with the magnitude and distribution of impacts from current and more stringent climate policies, and also with the effects of both implemented in concert.

Established emissions standards. In 2000, the Ministry of Environmental Protection issued GB 18352.1–2001, its first national standard on emissions from new road vehicles. Referred to as China 1 (for light-duty vehicles) and China I (for trucks and other heavy-duty vehicles), these specified quantities similar to the European Union's Euro 2/II (Directive 91/441/EEC and 91/542/EEC), or Euro 1/I, issued 8 years earlier.

China's national standards mandate the levels given in Table 2.2 on the following page for emissions from LDVs and HDVs, and in Table 2.3 on the next page for the presence of sulfur in fuels. Future national standards are specified, with final

Table 2.2: Recent Chinese tailpipe emissions standards, and selected European Union standards for comparison (ICCT 2014, and linked documents). Note 1: two quantities are given, for gasoline and diesel passenger cars respectively. Note 2: two quantites are given, for the European Static Cycle + European Load Response test; and the European Transient Cycle respectively, versions of which are specified by the Chinese standards.

Species	**CO**		**HC**		**HC+NO$_x$**		**NO$_x$**		**PM**		**NMHC**	
Light-duty passenger vehicles (g/km)												
China 3	2.3	0.64[n.1]	0.20	—	—	0.56	0.15	0.50	—	0.05	—	
China 4	1.0	0.50	0.10	—	—	0.30	0.08	0.25	—	0.025	—	
China 5	1.0	0.50	0.10	—	—	0.23	0.06	0.18	0.0045		—	
Euro 5	0.50		—		0.23		0.18		0.0045		—	
Euro 6	0.50		—		0.17		0.08		0.0045		—	
Heavy duty vehicles (g/kW·h)												
China III	2.1	5.45[n.2]	0.66	—	—		5.0		0.10	0.16	—	0.78
China IV	1.5	4.0	0.46	—	—		3.5		0.02	0.03	—	0.55
China V	1.5	4.0	0.46	—	—		2.0		0.02	0.03	—	0.55
Euro V	1.5	—	0.46		—		2.0		0.02		—	0.55
Euro VI	1.5	4.0	0.13		—		0.4		0.01		—	0.16

Table 2.3: Sulfur content, in parts per million, in established and future China fuel quality standards (ICCT 2014). Note 1: China I gasoline was required to be unleaded, but no maximum sulfur content was specified.

Level	I	II	III	IV	V
Gasoline	—[n.1]	500	150	50	10
Diesel	2000	500	350	50	10

implementation dates. Some provinces have sought approval to proceed with earlier implementation of these standards. This permission is partly predicated on the ability of fuel providers to supply cleaner fuels that will not degrade the emissions control technologies implied by the standards.

Finally, a variety of local, ad-hoc policies also aim to reduce emissions from road vehicles. These include prohibiting driving by some or all vehicles on certain days, accelerated retirement of older vehicles, limiting the number of vehicles owned, and promoting the adoption of New Energy Vehicles (alternative fuel vehicles, such as battery-electrics). The current analysis does not treat these policies, in effect assuming that their effect is constant as reflected in the transport activity and energy- and emissions intensities in the base year data.

Climate & energy policy. Climate and energy policies in the broader economy are another class of measures which can reduce emissions of the pollutants that contribute to poor air quality. Similar to transport-sector policy, these change the amount or type of energy used, or the amount of pollution emitted per unit energy. Section 2.4.3 describes in more detail how these changes contribute to reductions in total emissions.

Policy scenarios in this study. My analysis employs five model configurations, labeled A–E, as shown in Figure 2-1 on the following page.

The policies implemented in these scenarios are as follows.

A. No Policy. Pollutant emissions from all sectors, including transportation, remain the same per unit of fossil energy consumed, as they were in 2007 (the base year for the analysis). As energy demand grows in projections, associated pollutant emissions grow at the same rates. I also adopt the mild, autonomous reductions in energy-basis emissions factors in non-transport sectors developed by M. Li,

[1] Firms have a direct incentive to improve the efficiency of their production processes, thereby reducing costs. These improvements can have the side effect of improving energy efficiency or reducing pollution.

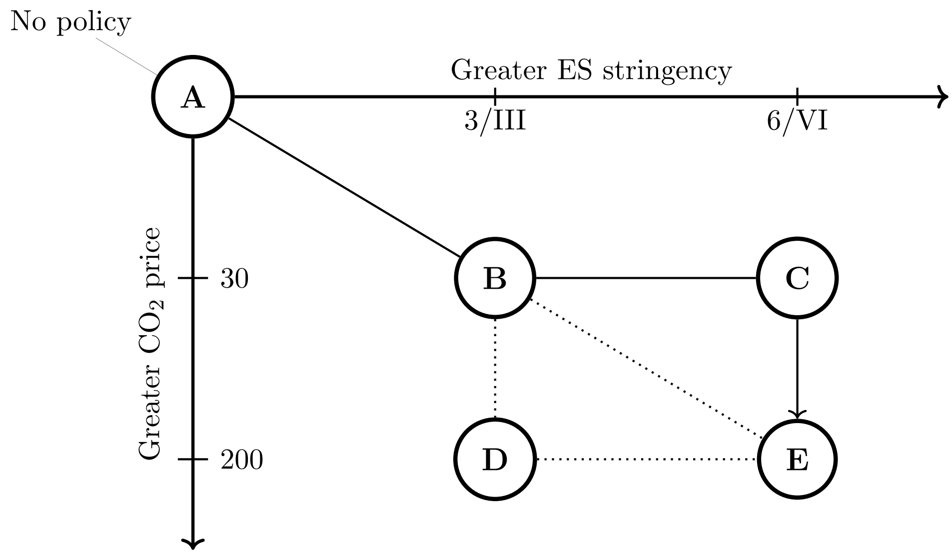


Figure 2-1: Policy scenarios in this study, with the stringency of road transport emissions standards and the initial (2015) CO_2 price level in RMB per ton.

N. Selin, et al. (2014), representing the impact of learning-by-doing[1] and capital turnover[2] (see Section 2.4.1).

I also consider a sensitivity scenario, termed **A′**, with no transportation ES implemented, and the small CO_2 price of Scenario B.

B. Established Policies. All new road vehicles and fuels meet the China 3/III standards, so that the entire fleet converges towards this standard over time as older, dirtier vehicles are retired. In regions which have already committed to introducing vehicles cleaner than China 3/III in the near future, the lower emissions levels are used instead. In addition, a small, gradually-rising, economy-wide CO_2 price promotes energy intensity improvement and fuel switching to reduce CO_2 emissions. This instrument is used to model the combined effect of China's prior and established national and regional energy- and carbon-intensity targets and other direct policy measures affecting the broader economy. As a result of the energy system changes induced by the CO_2 price, there is a co-benefit of

[2]Industrial equipment has a finite lifetime and must be periodically replaced. New, replacement equipment is often more efficient, requiring less energy or producing less emissions for the same production.

pollutant emissions reductions in these sectors (mainly as a result of displacing coal).

C. Stringent ES. More stringent tailpipe emissions and fuel quality standards are introduced, reaching China 6/VI nationwide starting from 2015. The CO_2 price is the same as in Scenario B.

D. Climate policy. A CO_2 price that is larger and rises more quickly, causing more rapid change in emissions across the entire economy. Road transport ES are the same as in Scenario B.

E. ES and climate policy. The combination of the stringent ES from Scenario C, and the higher CO_2 price from Scenario D.

Comparing Scenarios **A** and **B** illustrates how much established policies (in place prior to the introduction of the new nationwide China 4/IV standard) are expected to reduce pollutant emissions, compared to a future where transportation energy use has the same air pollutant emissions intensity as today. Comparing Scenario $\mathbf{A'}$ to **B** further isolates the effect of ES, as both scenarios have the same CO_2 price. Comparing Scenarios **B** and **C** similarly illustrates the impact of accelerating road transport policies under the same CO_2 price. Comparing Scenario **B** with Scenarios **C**, **D** and **E** illustrates the relative size and distribution of benefits from road transport policies compared to climate policy, and also the combined effect of the two.

2.3 Representing transport subsectors in the China Regional Energy Model

In order to represent policies that target road transport in particular, I extend the China Regional Energy Model (C-REM), a multi-sector, multi-region, recursive-dynamic CGE model of the global economy, with provincial detail in China. The model has 30 regions within China and four international regions (see Table 2.4 in

Appendix 2.A on page 75); the economy is represented in 14 sectors (see Table 2.5 on page 75). The C-REM projects output from each sector of each province, as well as trade and final demand (consumption), in value units, every 5 years to 2030.

Section 2.3.1 describes a disaggregation of the original model's monolithic transport sector into five subsectors; Section 2.3.2 discusses the data collected and used to calibrate the resulting structure.

2.3.1 Methodology

The transport sector disaggregation has two parts. *Commercial transport* denotes the activity of the original C-REM `TRN` sector. The portion of this sector's output consumed as an intermediate input in other production is assumed to be *freight transport*, while the portion directly consumed by households is *commercial passenger transport*. As shown in Figure 2-2 on page 48 (top panel), each of these portions is disaggregated into *road* and *non-road* sectors, with the latter comprising rail, marine, air and (for freight only) pipeline transport. *Household transport* denotes the consumption, by households, of commercial passenger transport (i.e. output from the former `TRN` sector), as well as from other sectors entering a household vehicle transport (`HVT`) sector with particular structure. In the CGE model's supplemental physical accounting, the activity levels for the two freight transport sectors—freight road (`FR`) and "other" i.e. non-road freight (`FO`)—are in t km/year, and the activity levels for the two commercial passenger sectors (passenger road (`PR`) and non-road passenger (`PO`)) and the `HVT` sectors are in pers. km/year.

The household transport structure (Figure 2-2 on page 48, bottom) is adapted from one used in the Economic Projection and Policy Analysis (EPPA) model, developed and detailed by Karplus et al. (2013) for the representation of light duty vehicle technology detail in CGE, and previously applied to China by Kishimoto, Paltsev, et al. (2012). For the present work, transport disaggregation is applied to Chinese provinces only, leaving an aggregated `TRN` sector in the international regions, and alternative powertrains are not represented, since they formed a small fraction of the household vehicle stock in 2007 (Gong et al. 2012).

2.3.2 Data for subsector disaggregation

Reliable transport data are important if models are to provide credible transport policy analyses, and statistics on China's economy, energy use and transport system present certain challenges. Guan et al. (2012) recorded a "gigaton gap" in reported CO_2 emissions—equivalent to Japan's national total—between a sum of provincial totals and the national figure given by the National Bureau of Statistics of China (NBSC). Provincial statistical bodies and the NBSC apply corrections for real or perceived misreporting or perform other undocumented adjustments to correct various errors (Holz 2004). These challenges were encountered during development of the C-REM. The strategy for assembling the model's social accounting matrix (SAM) and associated supplementary accounts (particularly of energy), described fully by D. Zhang, Karplus, et al. (2013) and D. Zhang, Rausch, et al. (2013), involved matching sums of provincial output totals to national totals or international data on trade flows, while preserving sectoral share information, which is consistent within provincial data sets.

In disaggregating the transport sector for this research, I produced a provincial data set by collating publicly-available information from NBSC yearbooks (China Communications & Transportation Association 2008, 2009, and earlier and subsequent editions from 1999 to 2012 inclusive), and examined it to identify necessary adjustments. While transport statistics reported by the NBSC do not show the kind of 'gap' identified by Guan et al.—i.e., national totals are consistent with the sum of corresponding provincial quantities—other statistical anomalies are found. In particular, year 2008 statistics show significant decreases relative to 2007 in marine freight

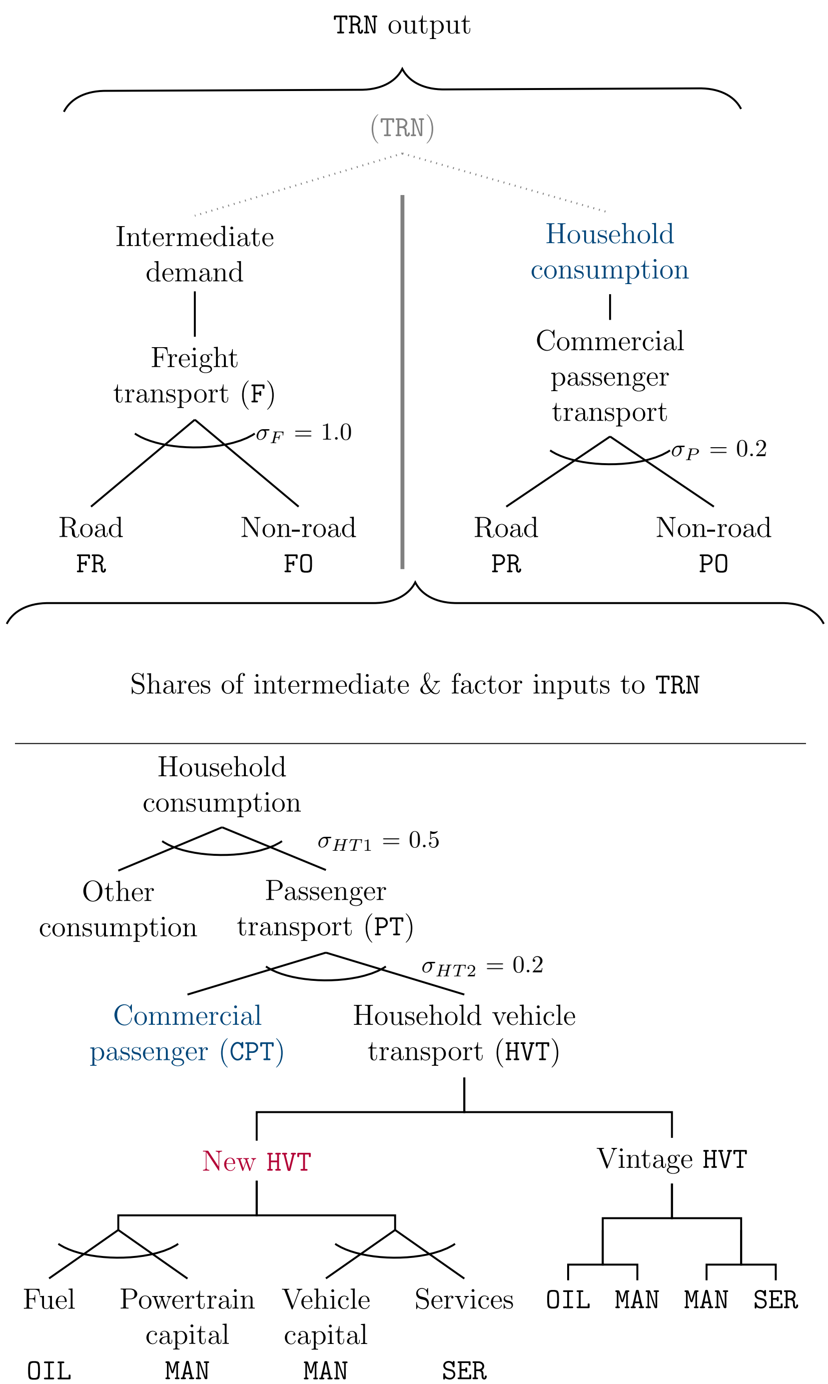


Figure 2-2: Disaggregate transportation representation in the C-REM. Top: a monolithic transportation services sector is broken into *freight* transport—supplying intermediate demand by other economic sectors—and *passenger* transport—supplying the demand of households for commercial travel. Bottom: households' consumption contains passenger transport, which consists of commercial passenger transport or own-supplied, household private vehicle transport. The latter is produced by households themselves, using fuel and vehicles; vehicle purchases consist of inputs from the manufacturing and service sectors.

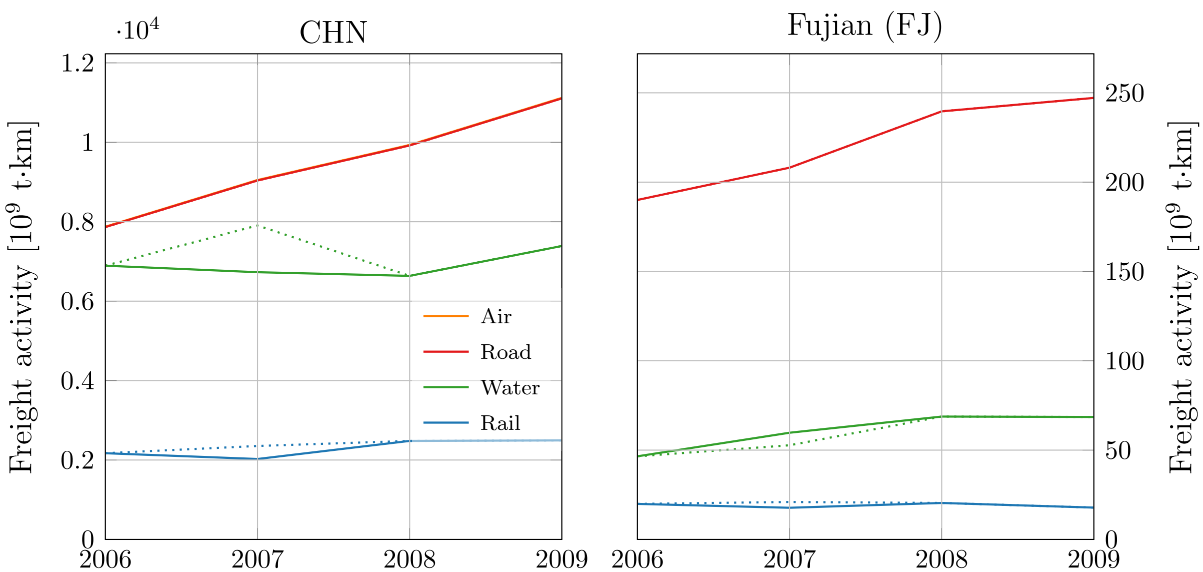


Figure 2-3: Stacked official (dotted) and adjusted (solid lines) data from the NBSC on freight activity for all of China (left) and an example province, Fujian (FJ, right). The official data, showing an apparent increase of more than 100% in national road freight t km (dotted green to solid red line) from 2007—2008, are evidence of an unpublished statistical correction.

transport activity (Figure 2-3).

There are alternate explanations for this abrupt change. One possibility is that the data accurately reflect a reduction in China's exports, due to falling global demand at the beginning of the 2008–2010 global recession. However, this theory fails to explain the corresponding increase in road freight activity, such that the growth in overall totals remained steady from 2007 through 2009 (Figure 2-3, solid lines). The growth rates of individual modes have also paralleled the national total in subsequent years. Instead, I interpret the change as a statistical adjustment—for a former over-reporting of marine traffic, under-reporting of highway traffic, or both—that was implemented suddenly in 2008, without notice or adjustment of previous data. In order to be consistent with the new reporting, I adjust the 2007 data by sharing out the total freight activity in the same proportion as in 2008, province by province (Figure 2-3, dotted lines). The resulting 2007 mode shares of freight activity are shown

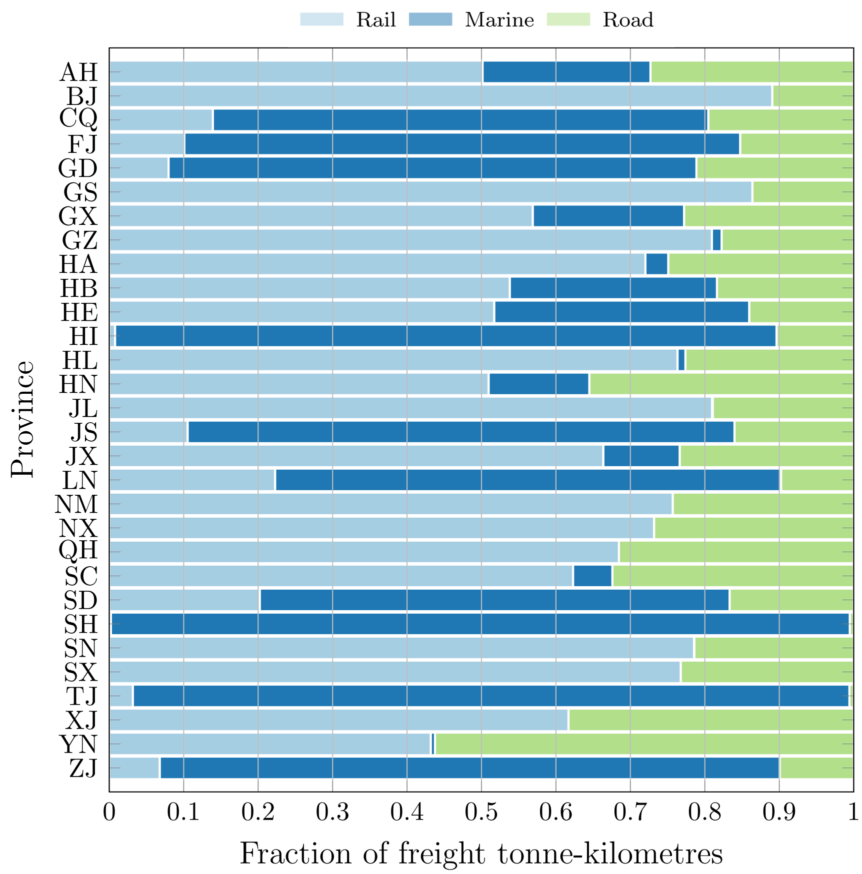


Figure 2-4: Adjusted mode shares of freight traffic, Chinese provinces, 2007. See Table 2.4 on page 75 for province names and locations. Note large variation in the share of road transport (FR) across provinces; and large share of the rail/marine split in non-road transport (FO).

in Figure 2-4, and reflect considerable province-to-province variation. For instance, shipping dominates in provinces with international (Shanghai, SH or Guangdong, GD) or domestic ports (Chongqing, CQ), while in other provinces the share of FR varies between 10% (Beijing, BJ) and 56% (Yunnan, YN) of the total.

While supplemental transport energy and physical activity accounts for country-level CGE regions can be based on non-governmental organization (NGO)- or privately-maintained international databases such as Passport GMID, IEA World Energy Balances and the International Road Federation's World Road Statistics (Euromonitor International 2011; International Energy Agency 2010; International Road Federation

2010), the China Energy Statistical Yearbook's provincial energy balance tables report total transport energy consumption by fuel; without any disaggregation by mode. I estimate the energy intensity of the four commercial transport sectors `FO`, `FR`, `PO` and `PR` by combining the adjusted activity levels from NBS with bottom-up energy intensity figures for each mode from the extensive literature on Chinese transport (see Kishimoto, X. Zhang, et al. 2013, Appendix A).

In order to divide the TRN output and input among the four sectors while maintaining a balanced SAM, `TRN` sector intermediate and factor demands are apportioned to the new sectors `FO`, `FR`, `PO` and `PR`, in proportion to their outputs. Energy demands (e.g. of `COL`, `CRU`, `ELE`, `GAS`, `GDT`, and `OIL`) are then rebalanced between `FO` and `FR` (likewise `PO` and `PR`) such that the relative energy intensities of the modes match our estimate from statistics and literature and the input of `OIL` to `FR` (likewise `PR`) reflects the 91–100% share (according to mode) of petroleum in Chinese road transport in 2007 (Ou et al. 2010).

2.4 Linking energy use to the REAS emissions inventory

2.4.1 Energy-basis emissions factors for transport subsectors

To represent the effects of the emissions policies discussed in Section 2.2.2, the physical accounts of the model were expanded to include primary pollutant species from the Regional Emissions in ASia (REAS) database, version 2.1 (Kurokawa et al. 2013): black carbon (BC), carbon monoxide (CO), nitrogen oxides (NO_X), organic carbon (OC), and sulfur dioxide (SO_2). Primary pollution is modeled as a byproduct of either combustion of fuels to produce energy, or of industrial or technical processes. I associate emissions of each species with individual sectors, provinces, and energy sources. This connection is made by calibrating energy-basis emissions factors (EFs), in mass of pollutant emitted per unit fuel energy consumed, using the base-year (2007) energy data contained in the C-REM SAM:

$$\text{Emissions factor}_{p,f,i,r} = \frac{\text{Emissions of } p_{f,i,r}}{\text{Consumption of } f_{i,r}} \quad \text{for every} \quad \begin{cases} \text{Pollutant} & p \\ \text{Fossil fuel} & f \\ \text{End-use sector} & i \\ \text{Province} & r \end{cases} \tag{2.1}$$

In the model projection, the product of an emissions factor and the C-REM projected demand for energy gives the quantity of emissions for each p, f, i, r.

The base-year (2007) data in the REAS v2.1 emissions inventory and the C-REM SAM imply EFs in the following way. For each province and species, I aggregate emissions from the REAS combustion and non-combustion sectors to C-REM sectors and REAS fuels to C-REM fuels. I divide these emissions totals by the corresponding energy flow from the C-REM supplemental accounts. The resulting 2007 EFs thus exactly reproduce the 2007 REAS v2.1 emissions totals when used with the C-REM base energy data.

For the C-REM road transport subsectors—`FR`, `PR`, and `HVT`—I determine subsector-specific EFs, and apply these to the C-REM projection of road transport fuel energy consumption. In the future, I make use of the detailed, bottom-up, fleet model of Akerlind (2013). This model tracks total Chinese vehicles in detailed categories by year of manufacture, representing the scrappage (conversely: survival) rate of older vehicles; improving fuel economy; and annual driving distance differences between newer and older vehicles. The fleet model also accounts for fuel demand using these highly disaggregate categories; newer vehicles' driving activity is associated with a greater fuel economy.

In C-REM periods beyond the base year, I use the engineering model to determine the portion of fuel energy demand attributable to vehicles which are 'new' since the prior C-REM period. For instance, for the C-REM forecast year 2010, this is the sum of fuel energy demand, in 2010, by vehicles sold in 2010, 2009, or 2008. The remainder of fuel energy demand in 2010 is associated with vehicles sold in 2007 or earlier. Using the engineering model in this way increases the time resolution in description of the vehicle fleet to individual years, allowing a more precise division of road transport energy use between specific numbers of new, standards-compliant vehicles, and older, higher-polluting vehicles. However, the internal C-REM vintaging logic, which expresses this split in economic value terms, is not constrained to be consistent with the engineering model outputs. The size of any mismatch is not measured or addressed in the current work.

Fuel demand from new vehicles is associated with EFs in Figure 2-5 on the next page (bold, horizontal lines). The remaining fuel demand, associated with pre-existing vehicles, retains the EF of the previous period—in 2010, this is the 2007 REAS/C-REM implied EF, or in 2015 or later, the energy-weighted average across new and used vehicles in the previous period. Thus, policy which reduces emissions in new vehicles relative to Scenario A also reduces the emissions associated with the fuel demand of vintage (used) vehicles in subsequent C-REM periods.

In Scenario B, new vehicles meet the China 3 (passenger road and household vehicle transport) or China III (freight road) standard from 2008 onwards. In Scenarios C and E, new vehicles meet China 3/III from 2007–2010, China 4/IV (etc.) from 2011–2015, and China 6/VI from 2016 onwards, excepting Beijing, which meets China 5/V in 2013 and China 6/VI in 2016.

China's emissions standards, like the Euro standards on which they are based,[3]

[3]e.g., Directives 91/441/EEC and 91/542/EEC, for Euro I/1 respectively.

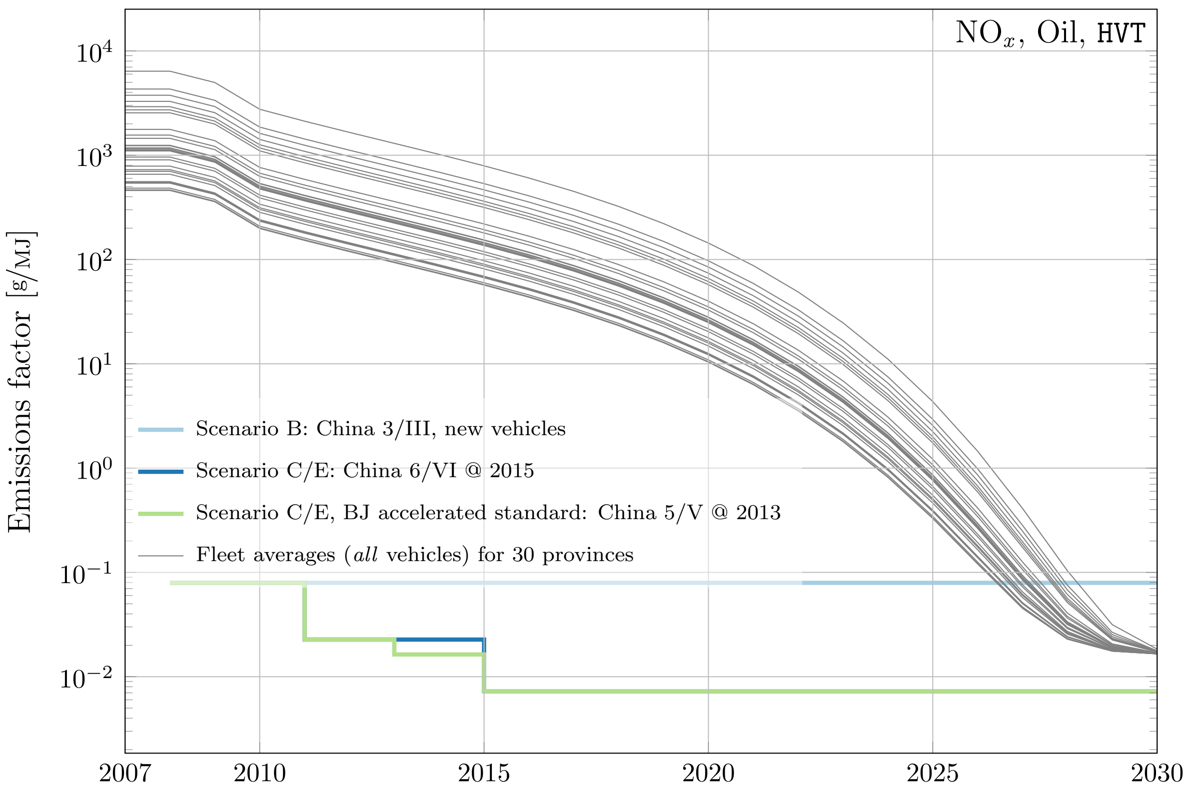


Figure 2-5: Bold lines: mandated emissions factors of NO_X from refined oil (gasoline or diesel) combustion in *new*, private, light-duty vehicles under baselines and policy scenarios. Thin lines: fleet-average `OIL` NO_X emissions factors for 30 provinces under Scenario C.

specify distance-basis, rather than energy-basis, EFs for new vehicles. To determine energy-basis EFs for the calculation just described, I use the on-road measurements of K. He, Yao, et al. (2010). For future Chinese standards (China 5/V and 6/VI), I assume the on-road emissions levels will be in the same proportion to China 4/IV as the regulated levels are to the China 4/IV regulated levels. For `PR` and `HVT`, I use the figures for light-duty gasoline (passenger) vehicles, and to represent the average `FR` vehicle, I use the figures for medium-duty diesel trucks.

Figure 2-5 on page 54 also shows the energy-weighted average EF across the entire fleet, for NO_X from refined oil combustion in the `HVT` fleet. Table A.1 in Appendix A gives a complete list across provinces for 2010, 2015 and 2020. In Scenario A, the implied 2007 EF is used unchanged throughout the projection.

Because there is a base-year implied EF for each species and province, the relative improvement in EF due to the introduction of lower-emission vehicles differs province-to-province, and species-to-species. Absent any differences in policy across provinces, EFs would eventually converge to the same level in all provinces, as today's heterogeneous provincial vehicle stocks are scrapped and replaced by vehicles with identical emissions characteristics; in the current projections, this occurs near the very end of the C-REM forecast period—see Figure 2-6.

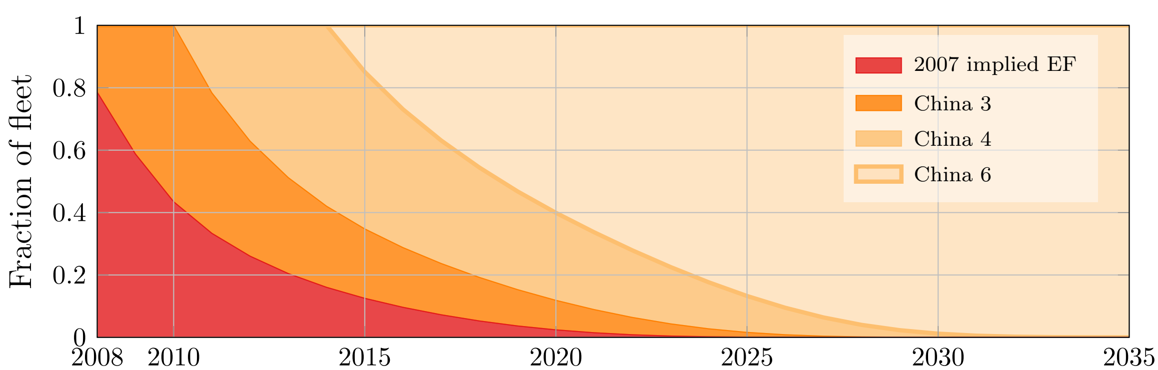


Figure 2-6: Projected fraction of private vehicle stock complying with various ES, Scenario C/D.

2.4.2 Emissions of non-transport sectors

Energy-basis emissions factors for all sources are applied to the non-transportation sectors to obtain a complete picture of economy-wide emissions, as described by M. Li, N. Selin, et al. (2014). Model base year (2007) EFs are calibrated, as in transportation, to reflect the total energy demand in the C-REM SAM and quantities in the REAS database. EFs undergo an exogenous, exponential decline, calibrated to reflect observations in 2010 and 2013, using the method of Webster et al. (2008).

The exogenous decline in these EFs represents the continuing effect of non-market policies and actions by firms which—for instance—will retire older equipment and replace it with new equipment which produce less emissions in operation (capital turnover); or implement efficiency improvements in production processes that also reduce pollution intensity (learning-by-doing). These trends are assumed to be independent of any CO_2 price applied to fossil fuel use in these non-transport sectors; the CO_2 price reduces emissions not by altering EFs, but by incentivizing low-emissions-intensity activities.

2.4.3 Policy impacts: modeled mechanisms and effects

The foregoing additions result in a model framework that can capture policy-related changes at the provincial and subsectoral level.

Economy-wide climate & energy policy. Climate policy in a CGE model such as C-REM signals sectors and households via changes in energy prices in proportion to CO_2 content, prompting these actors to respond with energy intensity improvements and input substitution to reduce CO_2 emissions. This economic response can include reductions in energy demand and switching to low carbon fuels, which may also reduce pollution in addition to CO_2.

Climate policy also has indirect impacts on the road transport sectors in two ways. First, freight transport demand arises in the economy because other economic sectors need to move their raw materials or finished goods to and from markets. Because a climate policy may cause each sector to increase or decrease production, their freight transport demands will also change, affecting the overall level of freight transport energy consumption and pollution. Second, households use their income to purchase passenger transportation services, private vehicles, and fuel. In a CGE framework, changes in household income mean more (due to economic growth), or less (due to stringent policy) income is available for these purchases. This, in turn, affects passenger transport demand, energy use, and pollution.

Road transport emissions standards. I model policy measures specifically aimed at reducing EFs more rapidly than they would decrease in the absence of regulation—in particular, road transport fuel quality standards and tailpipe emissions standards. The implied base-year EFs displayed in Table A.1 vary widely—by an order of magnitude for BC, CO and NO_X from household vehicles—reflecting province-to-province variation in the emissions attributed to road vehicles (in the REAS inventory), and the amount of energy used in household and commercial road transportation (as reflected in the official energy data underlying C-REM).

As a result, the relative improvement in EF due to the introduction of lower-emission vehicles in policy scenarios will differ province-to-province, and species-to-species. For instance, the 2010 EFs for black carbon (BC) are 11.5 and 25.9 g/MJ from road freight (`FR`) and household vehicle transport (`HVT`) in Anhui (AH) province, but 12.9 and 103 g/MJ in Gansu (GS) province. This suggests that implementing the same emissions standards will—absent differences in activity growth and mode share—produce a larger percentage decrease in `HVT` emissions in Gansu than in Anhui. On the other hand, because the EFs for road freight are the same, emissions standards will have a similar relative effect in these two provinces. The effect on total road transport emissions in these two provinces will in turn depend on the shares of these two modes, alongside passenger road, in total road transport energy use.

2.5 Results

I present findings in three sections: first, the impacts of emissions standards at different levels (Section 2.5.1); second, the way in which these impacts vary across provinces (Section 2.5.2), and third, the comparative impacts of economy-wide climate policy and emissions standards (Section 2.5.3).

2.5.1 Large reductions from road transport emissions standards

Despite average annual growth of 7.5% in transportation energy demand (4.5–9.9% across provinces) between 2010 and 2030, established polices reduce total national road transport pollution emissions to between 2% (OC) and 0.04% (CO) of their 2007 levels (Scenario B vs. Scenario A) (Figure 2-7 on the facing page). Further reductions occur in Scenarios C and E, as shown in Figure 2-8 on page 60.

If fully implemented, established policies will do most of the work; the continued sale of China 3/III vehicles alone will significantly reduce emissions in 2030, compared to the mix of vehicles currently in use. Although future standards (China 6/VI) reduce EFs by further orders of magnitude (cf. Figure 2-5 on page 54), these translate to a smaller absolute reduction in road transport emissions, because they act on a small base. These are, however, not trivial: in other countries, where large industrial sources of air pollutants have already been thoroughly regulated, these same absolute reductions in emissions from road sources would represent significant opportunities to reduce the health effects of poor air quality.

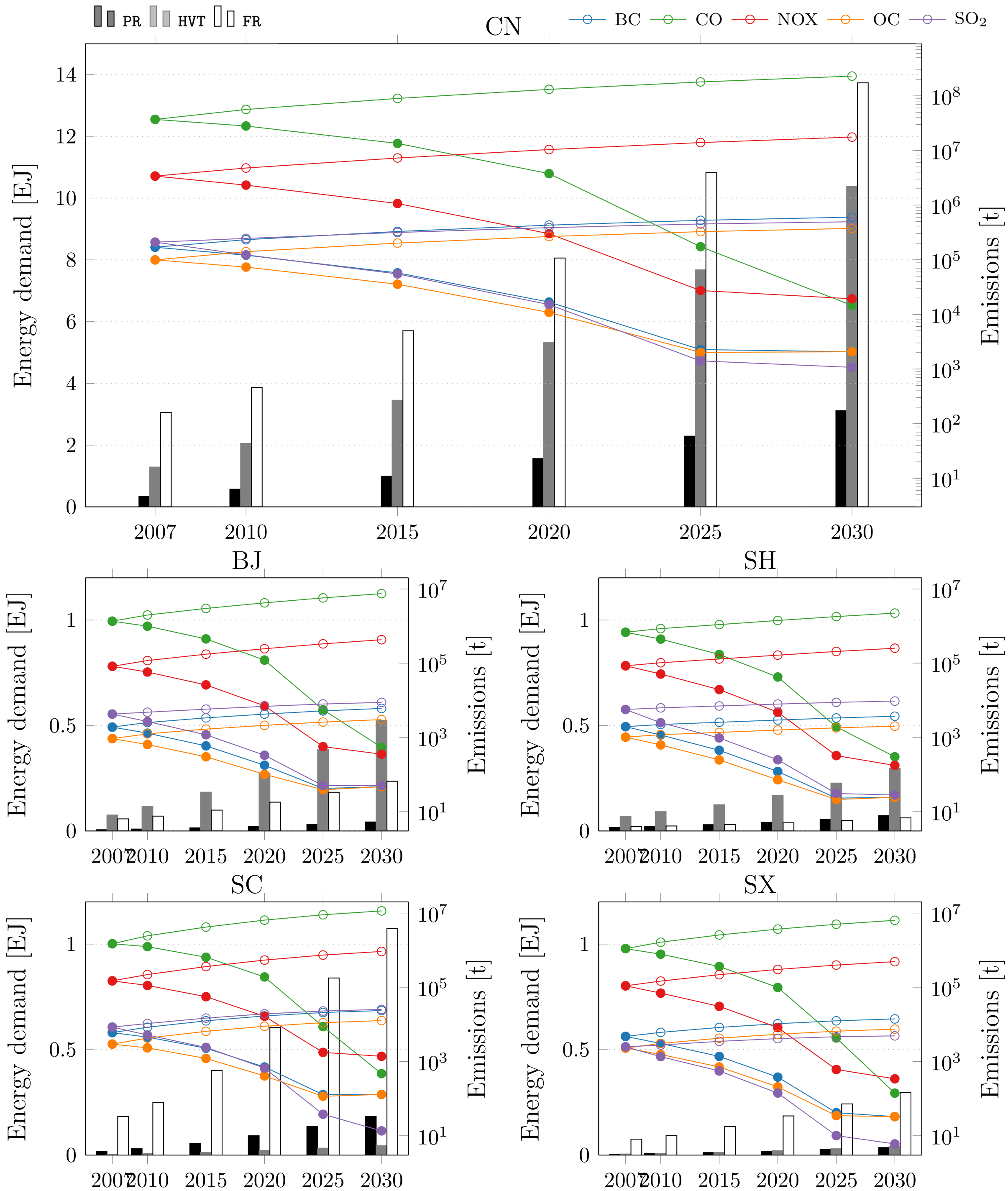


Figure 2-7: Bars and left ordinate: energy demand for three road transport sectors: commercial passenger (PR), household vehicles (HVT) and freight (FR). Lines and right ordinate: total emissions for five species in Scenario A (open marks) and Scenario B (filled marks). Top: China (CN) total; bottom: four selected provinces with distinct mixes of household vehicle, commercial passenger, and freight road transport—Beijing (BJ), Shanghai (SH), Sichuan (SC) and Shanxi (SX).

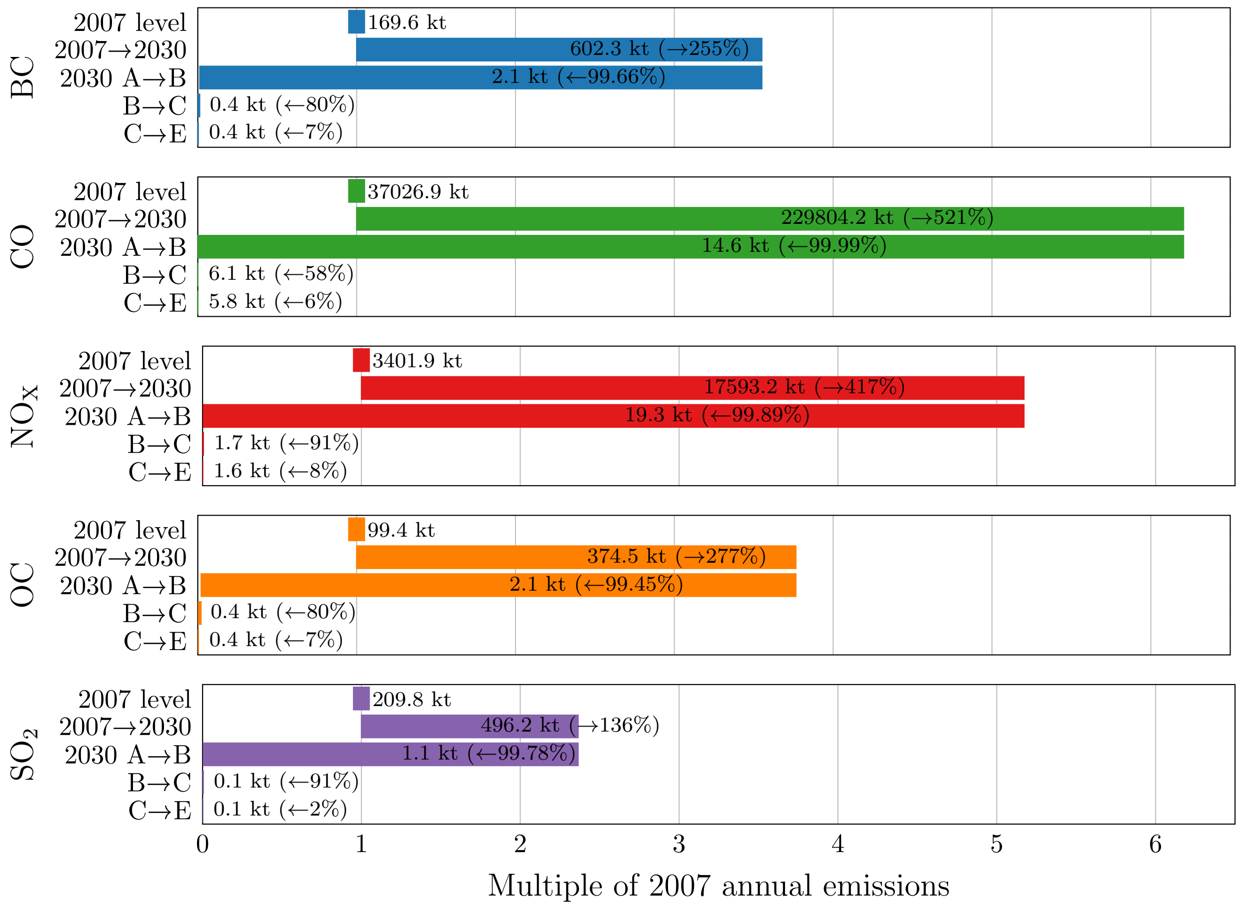


Figure 2-8: Changes in road transport emissions of five species: increase from 2007–2030 in Scenario A (no policy); and reductions in 2030 from Scenario A–B (introducing China 3/III and a mild CO_2 price), B–C (increasing ES stringency to China 6/VI), and C–E (increasing CO_2 price). The 2007 level is also shown, for reference. Annotations give the total emissions in each scenario and percent change in each year/scenario compared to the bar above.

2.5.2 Emissions standards impacts differ across provincial transport systems

Across China, road freight transportation is a larger consumer of energy (3.06 EJ in 2007) compared to the combination of private and commercial passenger transport (1.63 EJ in 2007)—and it is also a larger contributor of pollution: 0.6–6.9% of the national total in species besides CO, versus 0.03–6.8% for other road modes. Consequently, most pollution reductions due to ES occur in road freight transport—see Figure 2-11 in Appendix 2.A on page 75.

Figures 2-7 and 2-11 also illustrate that the extent of emissions reductions differs between provinces where passenger road travel activity is relatively large, and those where passenger road activity is small in comparison to freight. For instance, passenger road transport accounts for three quarters of road transport black carbon (BC) in Beijing, but only 36% in Chongqing.

2.5.3 Emissions standards are complementary to economy-wide climate and energy policy

Figure 2-9 on the following page illustrates that stringent road transport ES cause very modest additional reductions in total emissions of pollutants, even though they are very effective in reducing emissions *within* road transport sectors. For instance, China 6/VI emissions standards reduce road transport OC emissions by about 80% versus China 3/III, while more stringent climate policy reduces the same emissions by only 2–22% across provinces. This contrast is due to the small share of transport in overall emissions. In comparison, the mild CO_2 price of Scenario B causes 9.6–48% reductions in total emissions of pollution, versus no policy (Scenario A); similarly, tightening the CO_2 price only (Scenario D) results in 8.9–27% reduction versus established policy (Scenario B). Indeed, the co-benefits of climate policy for air pollution reduction are substantial, even for a relatively modest CO_2 price. As discussed above, co-

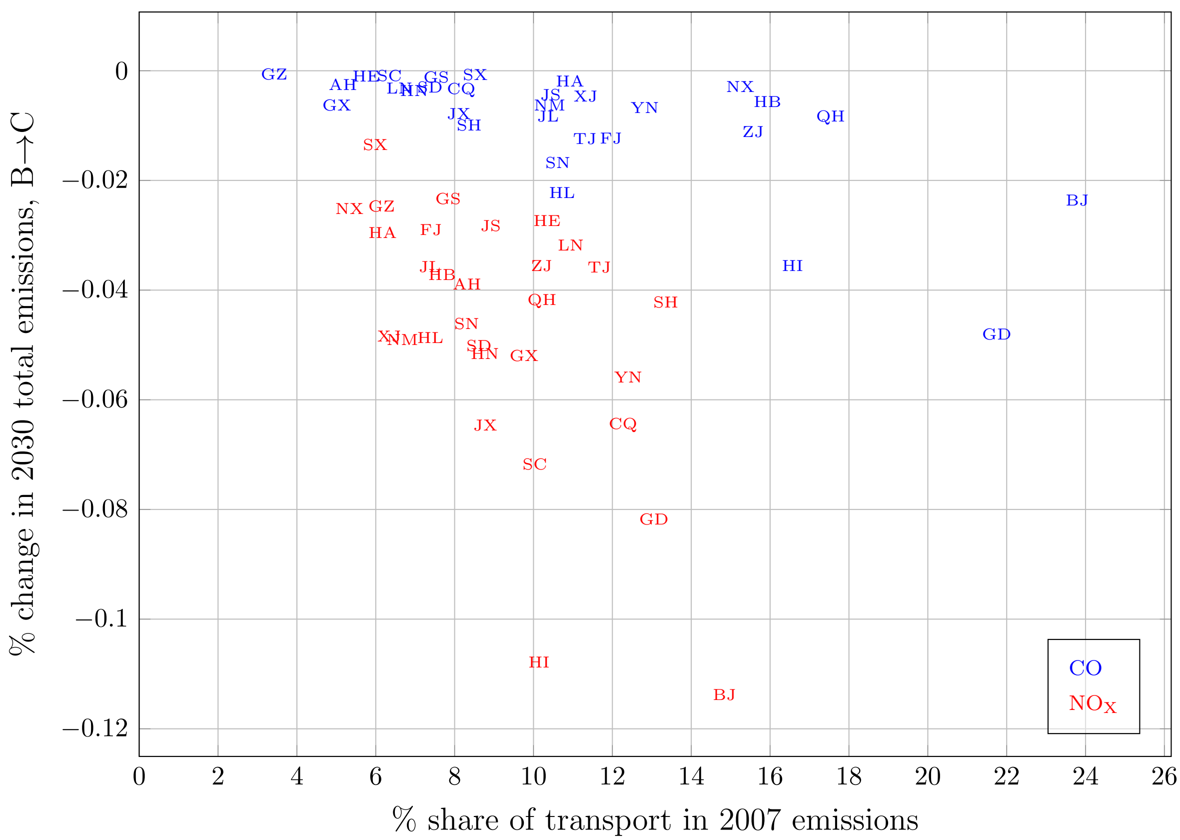


Figure 2-9: Fraction of transport CO and NO_X in *total* emissions in each province, versus the change in total CO and NO_X emissions due to moving from established to stringent ES. The reduction is generally smaller than 0.15%, in part because less than 24% or 15% of these species, respectively, is attributed to transport; and less to road transport.

benefits largely come from non-transport sectors, so emissions standards for road transportation are highly complementary in reducing pollution emissions.

Previous research emphasizes that the marginal cost of CO_2 emissions abatement in transportation tends to be higher than in other sectors, such as electricity and industry (Kishimoto, D. Zhang, et al. 2015). This means that responses to CO_2 pricing—efficiency improvements and fuel-switching—are smaller in transport, and the sectoral pollution co-benefit of CO_2 policy is also small. Indeed, reductions in air pollutant emissions due to CO_2 pricing in my scenarios mostly occur outside the transport sector: although increasing the CO_2 price (Scenario B→D) results in 8.9–27% additional reductions in total emissions, road transport emissions decrease by only 2.0–7.1% across species.

In contrast, the within-sector reductions due ES are measured in orders of magnitude, when comparing established policies (China 3/III) to a counterfactual future where the road vehicle fleet retained its 2007 emissions characteristics. Additional reductions due to tightening ES are similarly large as a percentage of remaining road transport emissions, yet small in absolute terms when compared to the co-benefit of economy-wide CO_2 pricing.

Consequently, transport-sector ES are an important complement to economy-wide climate policy, since they can achieve deep reductions via technology and cleaner fuel, which together greatly reduce EFs. To achieve the same transport-sector reductions purely through co-benefits of climate policy would require CO_2 prices much higher than the those modeled.

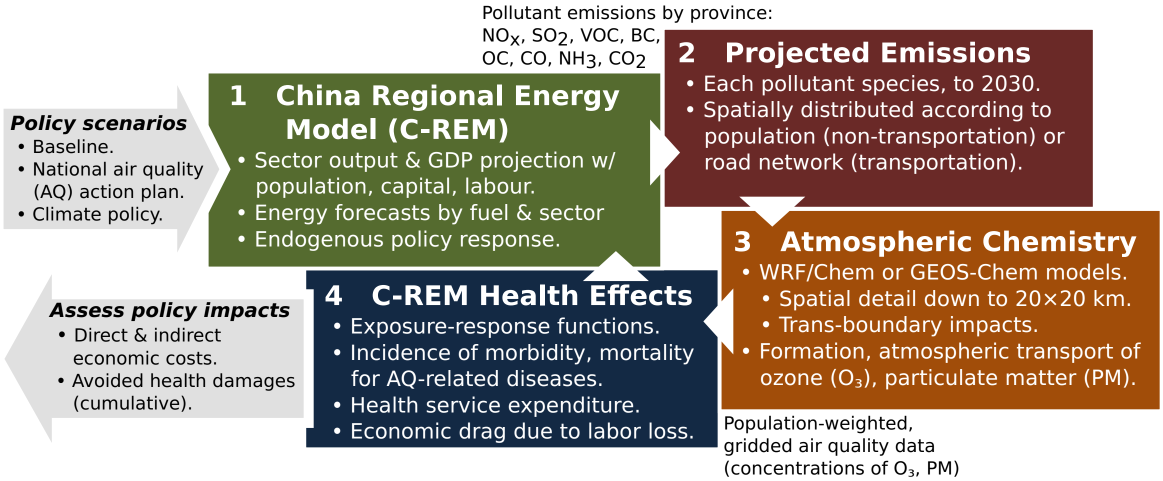


Figure 2-10: Regional Emissions, Air Quality, Consumption & Health (REACH) framework.

2.6 Discussion

2.6.1 Supporting atmospheric modeling and benefits quantification

M. Li, D. Zhang, et al. (2018) describe a Regional Emissions, Air Quality, Climate and Health (REACH) integrated assessment framework (Figure 2-10) for quantifying the human health impacts of Chinese air quality policies, and present results for varying CO_2 price scenarios. From the current research, the new, province-, sector-, and species-specific projections of road transport emissions can be used to support similar analysis. Quantified reductions or avoided primary pollution (steps 1 and 2, as in this work) forms input to atmospheric simulation (step 3), affecting chemistry and dynamics, leading to changes in the concentration of secondary pollutants which reflect the spatial movement of species. In step 4, reductions in population-weighted exposure to $PM_{2.5}$ and O_3 are translated to reductions in adverse human health effects, using future population densities and China-specific exposure-response relationships, in order to determine the change, due to policy, in the economic burden of pollution.

[4]n.b. this ‘transport’ is a term of art in atmospheric chemistry, referring to the vertical and lateral movement of pollutants with the air that contains them. This is distinct from ‘transport’ *qua*

One example of such use is Kishimoto, Karplus, et al. (2017), in which the regional chemical transport[4] model WRF-Chem version 3.5 was used to quantify the pollutant concentrations changes associated with the emissions changes identified here.[5] Figures 2-13 and 2-14 in Appendix 2.A illustrate the spatial effects of the emissions changes resulting from moving to China 3/III (Scenario B) from no China 3/III (Scenario A′). Figure 2-13 shows that reduction of BC, OC, CO, and NO_X emissions in 2020 mainly occurs in the North China Plain, Pearl River Delta, Yangtze River Delta, and Sichuan Basin. These emissions reductions would reduce the concentrations of $PM_{2.5}$ (8 µg/m^3 to 20 µg/m^3), NO_X (more than 20 µg/m^3), SO_2 (3 µg/m^3 to 5 µg/m^3) in Eastern China, and O_3 (12 µg/m^3 to 16 µg/m^3) in Southern China (Figure 2-14 on page 79). However, due to changes in the chemical dynamics of O_3 formation, the reductions would result in a slight *increase* in O_3 in the North China Plain. Kishimoto, Karplus, et al. (2017, Figure 9 on p.16) also show that no significant changes in air pollutant concentrations are observed when moving from China 3/III (Scenario B) to China 6/VI (Scenario C), in either 2020 or 2030, because road transport is a relatively small emissions source compared to other sources such as domestic energy uses and industry.

2.6.2 Policy implications and recommendations

Taken together, the results clearly illustrate the emissions reduction benefits of completely implementing emissions standards at the China 3/III level or higher. To quantify the consequences for secondary pollutant concentrations and the burden of human health impacts on the economy would require further work to exercise the complete REACH framework; but my results indicate the scale of the benefit available due to the two policy levers considered.

'transportation', used throughout the chapter.

[5] Full methodological detail and additional data sources are described at Kishimoto, Karplus, et al. 2017, p.8.

While moving to China 6/VI standards is clearly desirable, if tightening official stringency comes at the expense of sustained implementation effort that brings real-world fleets in line with policy targets, policymakers are better advised to focus on full enforcement of the existing standards. The results show that the marginal benefits of accelerating the policy timeline are modest. On the other hand, a small number of non-compliant vehicles, running on non-compliant fuels, could more than offset the modest benefits of moving a large number of sales to a more stringent standard; and indeed the work of X. Wu et al. (2016) and Yue et al. (2015) and others indicates ongoing issues with on-the-ground enforcement of existing standards.

Lessons from the longer history of air pollution policy Europe and elsewhere that suggest significant benefits from accelerated road vehicle standard implementation do not yet apply in China (Fenger 2009). Changes that result from incremental standard tightening are large relative to total emissions in today's European context, but remain small relative to total emissions in the Chinese context because of the still-large emissions from electric power and industrial emissions sources. As noted in Section 2.5.1 and Figure 2-7 on page 59, I project that energy demand for road transport will continue to grow through 2030, both in absolute terms and as a share of total energy demand, as increases in demand for transport more than offset improvements in energy efficiency (i.e., fuel economy). Consequently, road transport's share of total CO_2 emissions will also grow. In contrast, the large reduction in air pollutant emissions factors from implementing established (China 3/III) emissions standards means that the share of road transport in total air pollutant emissions will decrease markedly; and reductions from tightening ES can only further narrow this already-small share.

Therefore, I underscore that climate policies now being discussed and piloted, specifically a price on carbon such as the one I model, can serve as an important and effective complement to the full implementation of emissions standards in the road transportation sector. The work suggests two main policy recommendations.

First, policymakers can strengthen mechanisms for enforcing the newly-enacted China 4/IV emissions standards. Authorities at the highest levels should clearly direct the Standardization Administration of China and Ministry of Environmental Protection to strengthen and standardize the monitoring and enforcement system for fuel quality and vehicle tailpipe emissions. If the timing for increased stringency of standards—i.e., to China 6/VI levels—is defined well in advance, regulators, manufacturers and others in the system can adjust and prepare.

Second, continue to diligently work toward establishing a national CO_2 price with broad sectoral coverage. Although it seems likely that transportation will not be included in a national CO_2 emissions trading system, reductions in fossil energy use in other parts of the economy will deliver significant and meaningful co-benefits that will contribute to improved human health in the near to medium term.

2.6.3 Limitations: costs and implementation of tailpipe controls

A key difference between the carbon and ES policies implemented in these scenarios is that the cost of carbon emissions abatement is captured in the general equilibrium framework of C-REM. In contrast, the ES policies are implemented through exogenous calculation of road transportation EFs. Implementing tighter ES would require vehicle manufacturers to install more advanced ECT on passenger vehicles and road freight vehicles sold in China. Fuel economy improvements would also reduce the amount of energy used per kilometer, and thus contribute to reducing the amount of emissions per unit of fuel consumed.[6] Both of these compliance options impose costs on manufacturers and consumers if the resulting vehicles are more expensive than those that would be sold without the standard. Shao and Wagner (2015) estimate of

[6] On the other hand, ECT themselves have a small but measurable energy cost, reducing the fuel economy of vehicles, and again increasing the cost of compliance.

the costs of emissions control technologies for different levels of Euro standards, with some adjustment for the Chinese fleet, and note that the absolute costs for private, light-duty vehicles are nominal.

This study assumes that these increases in purchase price due to more advanced ECT, when considered as an increment on the cost of transport per passenger-kilometer or ton-kilometer, are not large enough to affect vehicle purchases or vehicle use intensity. More importantly, all scenarios assume that ES are *fully* implemented—that is, 100% of new vehicles comply with the active standard as of the sale date. In order to realize the air pollution emissions reductions identified here, ES implementation and compliance are critical; if a fraction of new road vehicles (either passenger or freight) are non-compliant, their much higher EFs will increase the fleet average so long as they remain in use. Figure 2-6 on page 55 shows that $\sim$ 40% of vehicles in 2020 have China 4/IV, 3/III, or pre-2007 EFs; as a consequence fleet-wide emissions that year under Scenarios B, C, or E (Figure 2-12 on page 77) are much higher than in 2030 (cf. Figure 2-8 on page 60), even though the 2020 fleets and total transport energy use are larger.

Future research could relax these assumptions to study, for instance, how the additional costs and efficiency penalties of road vehicle ECT and cleaner fuels affect activity growth and thus the benefits of ES. If reports are available of the degree of non-compliance with established standards, that information could be used to anticipate the erosion of the overall ES benefit of ES due to non-compliance—and thus the benefit of additional effort in enforcement.

2.6.4 Conclusions

Initial Chinese ES were based on previously established European Union regulations. Ever since, China's ES have been converging towards parity with the world's most stringent (ICCT 2014). Because transportation is one of several major polluting sectors, and because of the varied geographic distribution of transportation activity—and of air pollution and its health impacts—it is important to understand the contribution of transport-sector policies to improving air quality in China on a detailed, regional basis. It is also important to understand how coordination of multiple policies can lead to improved air quality in China, and whether this coordination is different from that required in other contexts.

To this end, I herein developed key components of an integrated assessment framework that projects the economic activities—including energy use in non-transport and transport sectors—giving rise to air pollution. Within this framework, I implemented two types of policies affecting emissions: road transport emissions standards, and economy-wide CO_2 pricing that gives a pollution co-benefit. Examining scenarios of no policy, established policies, and more stringent policies, I characterized the relative scale of their impacts on pollutant emissions within road transportation, and across the economy. My results indicate that increased CO_2 pricing, and full enforcement of more stringent road transport emissions standards, play complementary roles in reducing total emissions.

Acknowledgments

This chapter contains material first seen as Kishimoto, Karplus, et al. (2016) and Kishimoto, D. Zhang, et al. (2015) and some of which was later published in Kishimoto, Karplus, et al. (2017). This research builds on the work of the MIT-Tsinghua China Energy and Climate Project (CECP). I am grateful to my co-authors: Valerie J. Karplus and Da Zhang at MIT; Eri Saikawa and Min Zhong at Emory University, and Xiliang Zhang and Xu Zhang at Tsinghua University. Together we also acknowledge Prof. Zhang Qiang at Tsinghua University for ideas that helped shape the research, and Chiao-Ting Li, Audrey Resutek and Andrew J. Cockerill at MIT for helpful comments on the text. Any remaining errors or issues are my own.

References

Akerlind, Ingrid (2013). "Driving change: evaluating strategies to control automotive energy demand growth in China". S.M. thesis. Massachusetts Institute of Technology. HDL: 1721.1/81114 (↪ p. 52).

Aunan, Kristin, Jinghua Fang, Haakon Vennemo, Kenneth Oye, and Hans M. Seip (2004). "Co-benefits of climate policy—lessons learned from a study in Shanxi, China". In: *Energy Policy* 32.4. An economic analysis of climate policy: essays in honour of Andries Nentjes, pp. 567–581. ISSN: 0301-4215. DOI: 10.1016/S0301-4215(03)00156-3 (↪ p. 39).

Cheng, Shuiyuan, Jianlei Lang, Ying Zhou, Lihui Han, Gang Wang, and Dongsheng Chen (2013). "A new monitoring-simulation-source apportionment approach for investigating the vehicular emission contribution to the PM2.5 pollution in Beijing, China". In: *Atmospheric Environment* 79, pp. 308–316. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2013.06.043 (↪ p. 40).

China Communications & Transportation Association, ed. (2008). *Year Book of China Transportation & Communications*. 2008. Beijing, CN: Year Book House of China Transportation & Communications (↪ p. 47).

— ed. (2009). *Year Book of China Transportation & Communications*. 2009. Beijing, CN: Year Book House of China Transportation & Communications (↪ p. 47).

Deng, Xin (2006). "Economic costs of motor vehicle emissions in China: A case study". In: *Transportation Research Part D. Transport and Environment* 11.3, pp. 216–226. ISSN: 1361-9209. DOI: 10.1016/j.trd.2006.02.004 (↪ p. 39).

Euromonitor International (2011). *Passport Global Market Information Database*. Statistics retrieved 16 October 2011 (↪ p. 50).

Fenger, Jes (2009). "Air pollution in the last 50 years – From local to global". In: *Atmospheric Environment* 43.1. Atmospheric Environment - Fifty Years of Endeavour, pp. 13–22. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2008.09.061 (↪ p. 66).

Gong, Huiming, Michael Q. Wang, and Hewu Wang (2012). "New energy vehicles in China: policies, demonstration, and progress". In: *Mitigation and Adaptation Strategies for Global Change* 18.2, pp. 207–228. DOI: 10.1007/s11027-012-9358-6 (↪ p. 47).

Guan, Dabo, Zhu Liu, Yong Geng, Sören Lindner, and Klaus Hubacek (2012). "The gigatonne gap in China's carbon dioxide inventories". In: *Nature Climate Change* 2, pp. 672–675. DOI: 10.1038/nclimate1560 (↪ p. 47).

Guo, Xiurui, Liwei Fu, Muse Ji, Jianlei Lang, Dongsheng Chen, and Shuiyuan Cheng (2016). "Scenario analysis to vehicular emission reduction in Beijing-Tianjin-Hebei (BTH) region, China". In: *Environmental Pollution* 216, pp. 470–479. ISSN: 0269-7491. DOI: 10.1016/j.envpol.2016.05.082 (↪ p. 39).

Harmelen, Toon van, Joost Bakker, Bert de Vries, Detlef van Vuuren, Michel den Elzen, and Petra Mayerhofer (2002). "Long-term reductions in costs of controlling regional air pollution in Europe due to climate policy". In: *Environmental Science & Policy* 5.4. Linkages Between Regional Air Pollution and Climate Change in

Europe, pp. 349–365. ISSN: 1462-9011. DOI: 10.1016/S1462-9011(02)00043-6 (↪ p. 38).

He, Kebin, Y. Lei, X. Pan, Y. Zhang, Q. Zhang, and D. Chen (2010). "Co-benefits from energy policies in China". In: *Energy* 35.11. Energy and Its Sustainable Development for China, pp. 4265–4272. ISSN: 0360-5442. DOI: 10.1016/j.energy.2008.07.021 (↪ p. 39).

He, Kebin, Zhiliang Yao, and Yingzhi Zhang (2010). "Characteristics of Vehicle Emissions in China Based on Portable Emission Measurement System". In: *19th Annual International Emission Inventory Conference*. U.S. Environmental Protection Agency. San Antonio, TX, US. URL: http://www.epa.gov/ttnchie1/conference/ei19/session6/he.pdf (↪ p. 55).

He, Ling-Yun and Lu-Yi Qiu (2016). "Transport demand, harmful emissions, environment and health co-benefits in China". In: *Energy Policy* 97, pp. 267–275. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2016.07.037 (↪ p. 39).

Holz, Carsten A. (2004). "China's Statistical System in Transition: Challenges, Data Problems, and Institutional Innovations". In: *Review of Income and Wealth* 50.3, pp. 381–409. DOI: 10.1111/j.0034-6586.2004.00131.x (↪ pp. 47, 109, 165).

Hong, Chaopeng, Q. Zhang, K. He, D. Guan, M. Li, F. Liu, and B. Zheng (2016). "Variations of China's emission estimates response to uncertainties in energy statistics". In: *Atmospheric Chemistry and Physics Discussions* 2016, pp. 1–23. DOI: 10.5194/acp-2016-459 (↪ p. 40).

Hu, Jianlin, Li Wu, Bo Zheng, Qiang Zhang, Kebin He, Qing Chang, Xinghua Li, Fumo Yang, Qi Ying, and Hongliang Zhang (2015). "Source contributions and regional transport of primary particulate matter in China". In: *Environmental Pollution* 207, pp. 31–42. ISSN: 0269-7491. DOI: 10.1016/j.envpol.2015.08.037 (↪ p. 40).

Huang, Xiaoyan, Yang Wang, Zhenyu Xing, and Ke Du (2016). "Emission factors of air pollutants from CNG-gasoline bi-fuel vehicles: Part II. CO, HC and NOx". In: *Science of The Total Environment* 565, pp. 698–705. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2016.05.069 (↪ p. 40).

ICCT (2014). *China: Light-duty: Emissions*. URL: http://transportpolicy.net/index.php?title=China:_Light-duty:_Emissions (↪ pp. 42, 68).

International Energy Agency (2010). *World Energy Balances 1970-2009*. Accessed October, 2011. DOI: 10.1787/data-00512-en (↪ p. 50).

International Road Federation (2010). *World Road Statistics 2010 and electronic files* (↪ p. 50).

Karplus, Valerie J., Sergey Paltsev, Mustafa H. Babiker, and John M. Reilly (2013). "Applying engineering and fleet detail to represent passenger vehicle transport in a computable general equilibrium model". In: *Economic Modelling* 30, pp. 295–305. DOI: 10.1016/j.econmod.2012.08.019 (↪ pp. 46, 162, 196).

Kishimoto, Paul Natsuo, Valerie J. Karplus, Min Zhong, Eri Saikawa, Xu Zhang, and Xiliang Zhang (2016). "Impact of Coordinated Policies on Air Pollution Emissions from Road Transportation in China". In: 95th Annual Meeting of the TRB. (Washington, DC, US, Jan. 10, 2016). 16-6233. Transportation Research Board of the U.S. National Academies (↪ p. 69).

Kishimoto, Paul Natsuo, Valerie J. Karplus, Min Zhong, Eri Saikawa, Xu Zhang, and Xiliang Zhang (2017). "Impact of Coordinated Policies on Air Pollution Emissions from Road Transportation in China". In: *Transportation Research Part D. Transport and Environment* 54, pp. 30–49. ISSN: 1361-9209. DOI: 10.1016/j.trd.2017.02.012 (↪ pp. 65, 69).

Kishimoto, Paul Natsuo, Sergey Paltsev, and Valerie J. Karplus (2012). *The Future Energy and GHG Emissions Impact of Alternative Personal Transportation Pathways in China.* JP Report 231. Cambridge, MA, US: MIT Joint Program on the Science & Policy of Global Change. HDL: 1721.1/73607 (↪ pp. 46, 200).

Kishimoto, Paul Natsuo, Da Zhang, Xiliang Zhang, and Valerie J. Karplus (2015). "Modeling Regional Transportation Demand in China and Impacts of a National Carbon Policy". In: *Transportation Research Record. Journal of the Transportation Research Board* 2454, pp. 1–11. ISSN: 0361-1981. DOI: 10.3141/2454-01 (↪ pp. 63, 69).

Kishimoto, Paul Natsuo, Xiliang Zhang, Da Zhang, and Valerie J. Karplus (2013). "Modeling regional transportation demand in China and the impacts of a national carbon constraint". In: *16th Annual Conference on Global Economic Analysis.* Shanghai, CN (↪ p. 51).

Kurokawa, J., T. Ohara, T. Morikawa, S. Hanayama, G. Janssens-Maenhout, T. Fukui, K. Kawashima, and H. Akimoto (2013). "Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in ASia (REAS) version 2". In: *Atmospheric Chemistry and Physics* 13, pp. 11019–11058. DOI: 10.5194/acp-13-11019-2013 (↪ pp. 37, 40, 41, 51).

Lang, Jianlei, Shuiyuan Cheng, Wei Wei, Ying Zhou, Xiao Wei, and Dongsheng Chen (2012). "A study on the trends of vehicular emissions in the Beijing–Tianjin–Hebei (BTH) region, China". In: *Atmospheric Environment* 62, pp. 605–614. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2012.09.006 (↪ p. 39).

Li, Mingwei, Noelle Selin, Valerie J. Karplus, Chiao-Ting Li, Da Zhang, Xiaohu Luo, and Xiliang Zhang (2014). "Effects of China's Energy Policy on Future Air Quality in China and the U.S.". In: *2014 AGU Fall Meeting.* Session A24A-08. American Geophysical Union. San Francisco, CA, US (↪ pp. 43, 56).

Li, Mingwei, Da Zhang, Chiao-Ting Li, Kathleen M. Mulvaney, Noelle E. Selin, and Valerie J. Karplus (2018). "Air quality co-benefits of carbon pricing in China". In: *Nature Climate Change* 8.5, pp. 398–403. ISSN: 1758-6798. DOI: 10.1038/s41558-018-0139-4 (↪ p. 64).

Lu, Ming, Cong Sun, and Siqi Zheng (2017). "Congestion and pollution consequences of driving-to-school trips: A case study in Beijing". In: *Transportation Research Part D. Transport and Environment* 50, pp. 280–291. ISSN: 1361-9209. DOI: 10.1016/j.trd.2016.10.023 (↪ p. 39).

Miyazaki, Kazuyuki and Henk Eskes (2013). "Constraints on surface NOx emissions by assimilating satellite observations of multiple species". In: *Geophysical Research Letters* 40.17, pp. 4745–4750. ISSN: 1944-8007. DOI: 10.1002/grl.50894 (↪ p. 40).

Nam, Kyung-Min, Noelle E. Selin, John M. Reilly, and Sergey Paltsev (2010). "Measuring welfare loss caused by air pollution in Europe: A CGE analysis". In:

Energy Policy 38.9. Special Section on Carbon Emissions and Carbon Management in Cities with Regular Papers, pp. 5059–5071. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2010.04.034 (↪ p. 38).

Nam, Kyung-Min, Caleb J. Waugh, Sergey Paltsev, John M. Reilly, and Valerie J. Karplus (2013). "Carbon co-benefits of tighter SO_2 and NO_X regulations in China". In: *Global Environmental Change* 23.6, pp. 1648–1661. DOI: 10.1016/j.gloenvcha.2013.09.003 (↪ p. 39).

— (2014). "Synergy between pollution and carbon emissions control: Comparing China and the United States". In: *Energy Economics* 46, pp. 186–201. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2014.08.013 (↪ p. 39).

Ou, Xunmin, Xiliang Zhang, and Shiyan Chang (2010). "Scenario analysis on alternative fuel/vehicle for China's future road transport: Life-cycle energy demand and GHG emissions". In: *Energy Policy* 38.8, p. 3943. DOI: 10.1016/j.enpol.2010.03.018 (↪ p. 51).

Rive, Nathan (2010). "Climate policy in Western Europe and avoided costs of air pollution control". In: *Economic Modelling* 27.1, pp. 103–115. ISSN: 0264-9993. DOI: 10.1016/j.econmod.2009.07.025 (↪ p. 38).

Shao, Zhenying and David Vance Wagner (2015). *Costs and Benefits of Motor Vehicle Emission Control Programs in China*. Tech. rep. Washington, DC, US. URL: http://www.theicct.org/costs-and-benefits-motor-vehicle-emission-control-programs-china (↪ p. 67).

Wang, Haikun, Lixin Fu, Yu Zhou, Xuan Du, and Weihua Ge (2010). "Trends in vehicular emissions in China's mega cities from 1995 to 2005". In: *Environmental Pollution* 158.2, pp. 394–400. ISSN: 0269-7491. DOI: 10.1016/j.envpol.2009.09.002 (↪ p. 39).

Webster, Mort D., Sergey Paltsev, John Parsons, John M. Reilly, and Henry D. Jacoby (2008). *Uncertainty in Greenhouse Gas Emissions and Costs of Atmospheric Stabilization*. JP Report 165. Cambridge, MA, US: MIT Joint Program on the Science & Policy of Global Change. HDL: 1721.1/44623 (↪ p. 56).

Wu, Rongrong, Yu Bo, Jing Li, Lingyu Li, Yaqi Li, and Shaodong Xie (2016). "Method to establish the emission inventory of anthropogenic volatile organic compounds in China and its application in the period 2008–2012". In: *Atmospheric Environment* 127, pp. 244–254. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2015.12.015 (↪ p. 40).

Wu, Xiaomeng, Ye Wu, Shaojun Zhang, Huan Liu, Lixin Fu, and Jiming Hao (2016). "Assessment of vehicle emission programs in China during 1998–2013: Achievement, challenges and implications". In: *Environmental Pollution* 214, pp. 556–567. ISSN: 0269-7491. DOI: 10.1016/j.envpol.2016.04.042 (↪ pp. 39, 66).

Xia, Yinmin, Yu Zhao, and Chris P. Nielsen (2016). "Benefits of China's efforts in gaseous pollutant control indicated by the bottom-up emissions and satellite observations 2000–2014". In: *Atmospheric Environment* 136, pp. 43–53. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2016.04.013 (↪ p. 40).

Yang, Sheng and Ling-Yun He (2016). "Fuel demand, road transport pollution emissions and residents' health losses in the transitional China". In: *Transportation Research Part D. Transport and Environment* 42, pp. 45–59. ISSN: 1361-9209.

DOI: 10.1016/j.trd.2015.10.019 (↪ p. 39).

Yue, Xin, Ye Wu, Jiming Hao, Yuan Pang, Yao Ma, Yi Li, Boshi Li, and Xiaofeng Bao (2015). "Fuel quality management versus vehicle emission control in China, status quo and future perspectives". In: *Energy Policy* 79, pp. 87–98. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2015.01.009 (↪ pp. 39, 66).

Zhang, Da, Valerie J. Karplus, Sebastian Rausch, and Xiliang Zhang (2013). *Analyzing the Regional Impact of a Fossil Energy Cap in China*. Report 237. Cambridge, MA, US: MIT Joint Program on the Science and Policy of Global Change. URL: http://globalchange.mit.edu/research/publications/2376 (↪ p. 47).

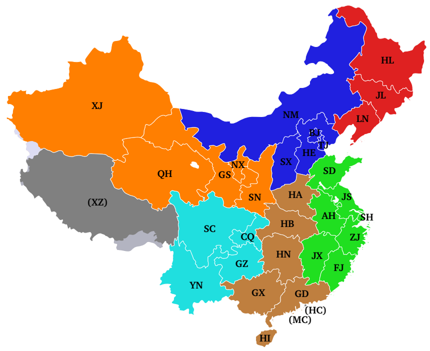
Zhang, Da, Sebastian Rausch, Valerie J. Karplus, and Xiliang Zhang (2013). "Quantifying regional economic impacts of CO_2 intensity targets in China". In: *Energy Economics* 40, pp. 687–701. DOI: 10.1016/j.eneco.2013.08.018 (↪ pp. 22, 47, 75).

Zhang, Shaojun, Ye Wu, Bin Zhao, Xiaomeng Wu, Jiawei Shu, and Jiming Hao (2016). "City-specific vehicle emission control strategies to achieve stringent emission reduction targets in China's Yangtze River Delta region". In: *Journal of Environmental Sciences*. ISSN: 1001-0742. DOI: 10.1016/j.jes.2016.06.038 (↪ p. 39).

Zheng, Xuan, Ye Wu, Jingkun Jiang, Shaojun Zhang, Huan Liu, Shaojie Song, Zhenhua Li, Xiaoxiao Fan, Lixin Fu, and Jiming Hao (2015). "Characteristics of On-road Diesel Vehicles: Black Carbon Emissions in Chinese Cities Based on Portable Emissions Measurement". In: *Environmental Science & Technology* 49.22. PMID: 26462141, pp. 13492–13500. DOI: 10.1021/acs.est.5b04129 (↪ p. 40).

2.A Additional tables and figures

Table 2.4: C-REM regions/Chinese provinces. Hong Kong (`HK`), Macau (`MC`) and Xizhang (Tibet, `XZ`) are not included in the C-REM; aggregate international regions not shown (D. Zhang, Rausch, et al. 2013).



Code	Name	Code	Name
`AH`	安徽 Anhui	`JS`	江苏 Jiangsu
`BJ`	北京 Beijing	`JX`	江西 Jiangxi
`CQ`	重庆 Chongqing	`LN`	辽宁 Liaoning
`FJ`	福建 Fujian	`NM`	内蒙古 Inner Mongolia
`GD`	广东 Guangdong	`NX`	宁夏 Ningxia
`GS`	甘肃 Gansu	`QH`	青海 Qinghai
`GX`	广西 Guangxi	`SC`	四川 Sichuan
`GZ`	贵州 Guizhou	`SD`	山东 Shandong
`HA`	河南 Henan	`SH`	上海 Shanghai
`HB`	湖北 Hubei	`SN`	陕西 Shaanxi
`HE`	河北 Hebei	`SX`	山西 Shanxi
`HI`	海南 Hainan	`TJ`	天津 Tianjin
`HL`	黑龙江 Heilongjiang	`XJ`	新疆 Xinjiang
`HN`	湖南 Hunan	`YN`	云南 Yunnan
`JL`	吉林 Jilin	`ZJ`	浙江 Zhejiang

Table 2.5: List of C-REM sectors, omitting the transportation subsectors shown in Figure 2-2 on page 48 (D. Zhang, Rausch, et al. 2013).

Code	Sector	Code	Sector
`AGR`	Agriculture	`MAN`	Other manufacturing industries
`COL`	Coal mining & processing	`OIL`	Petroleum refining, coking and fuels
`CON`	Construction	`OMN`	Metal, minerals, other mining
`CRU`	Crude petroleum products	`SER`	Services
`EIS`	Energy-intensive industries	`TRN`	Transportation & post
`ELE`	Electricity & heat	`WTR`	Water
`GAS`	Natural gas products	`c`, `g`, `i`	Final demands

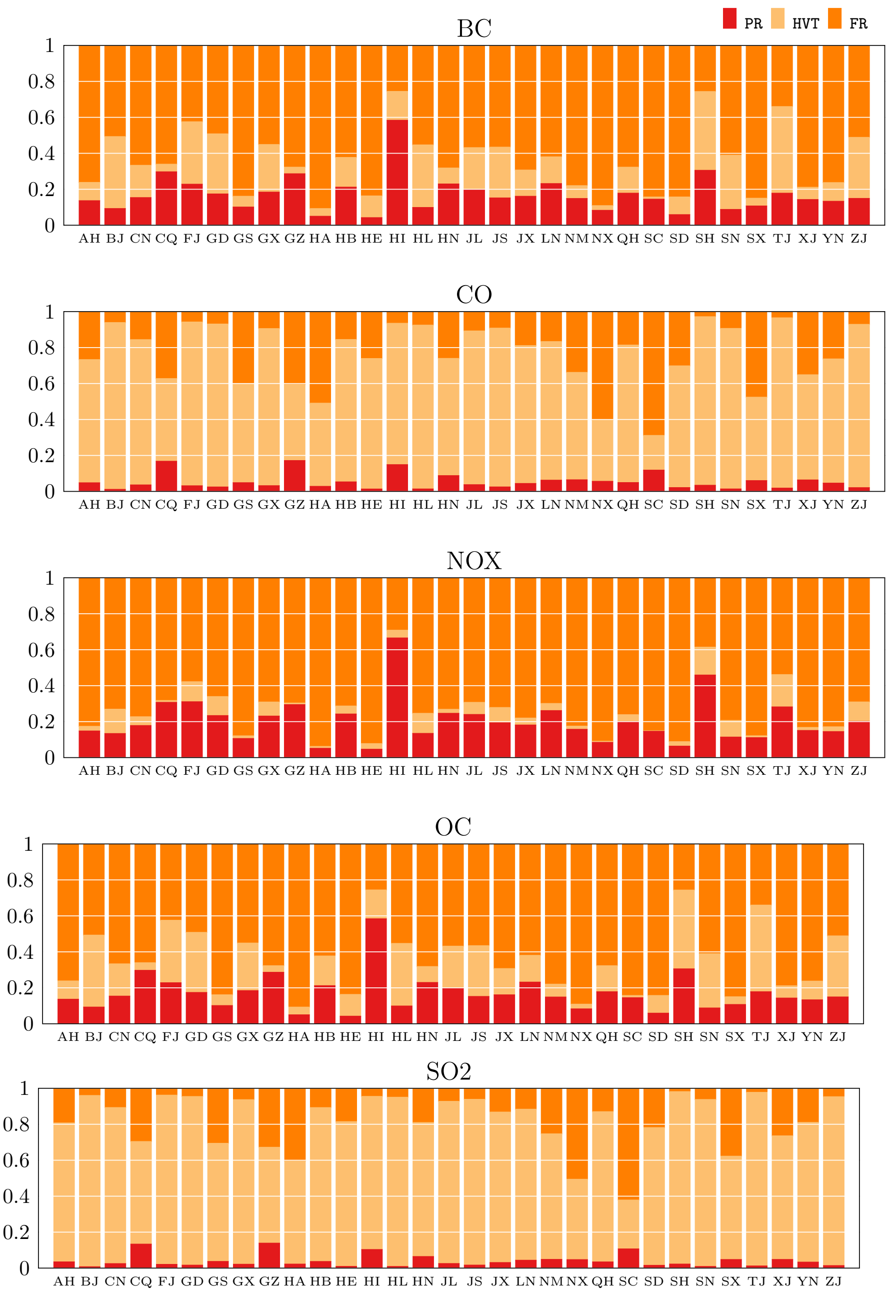


Figure 2-11: Contribution of each road transport mode to the total reduction of road transport emissions in 2030 due to stringent ES (Scenario B→C), by province.

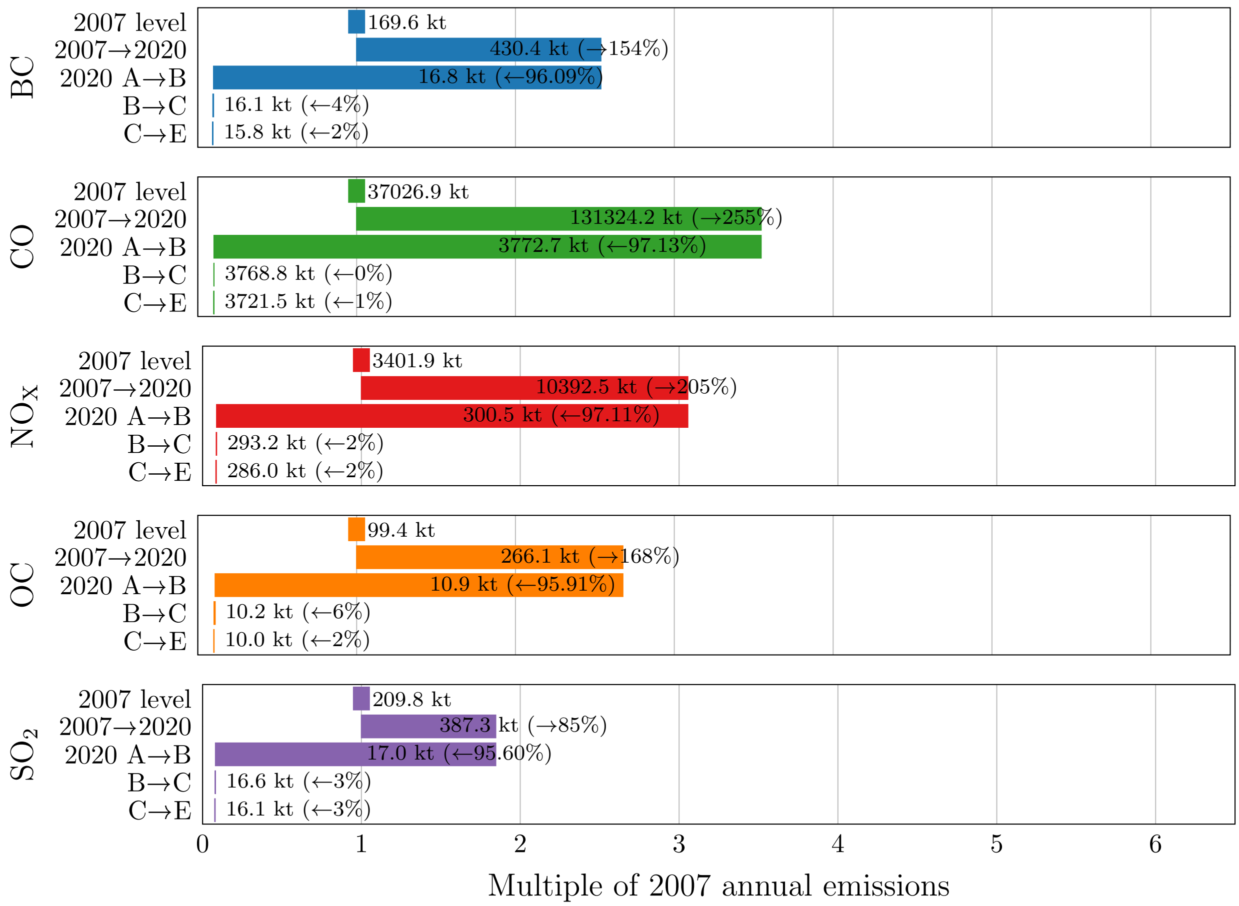


Figure 2-12: Changes in road transport emissions of five species: increase from 2007–2020 in Scenario A (no policy); and reductions in 2020 from Scenario A–B (introducing China 3/III and a mild CO_2 price), B–C (increasing ES stringency to China 6/VI), and C–E (increasing CO_2 price). The 2007 level is also shown, for reference. Annotations give the total emissions in each scenario and percent change in each year/scenario compared to the bar above.

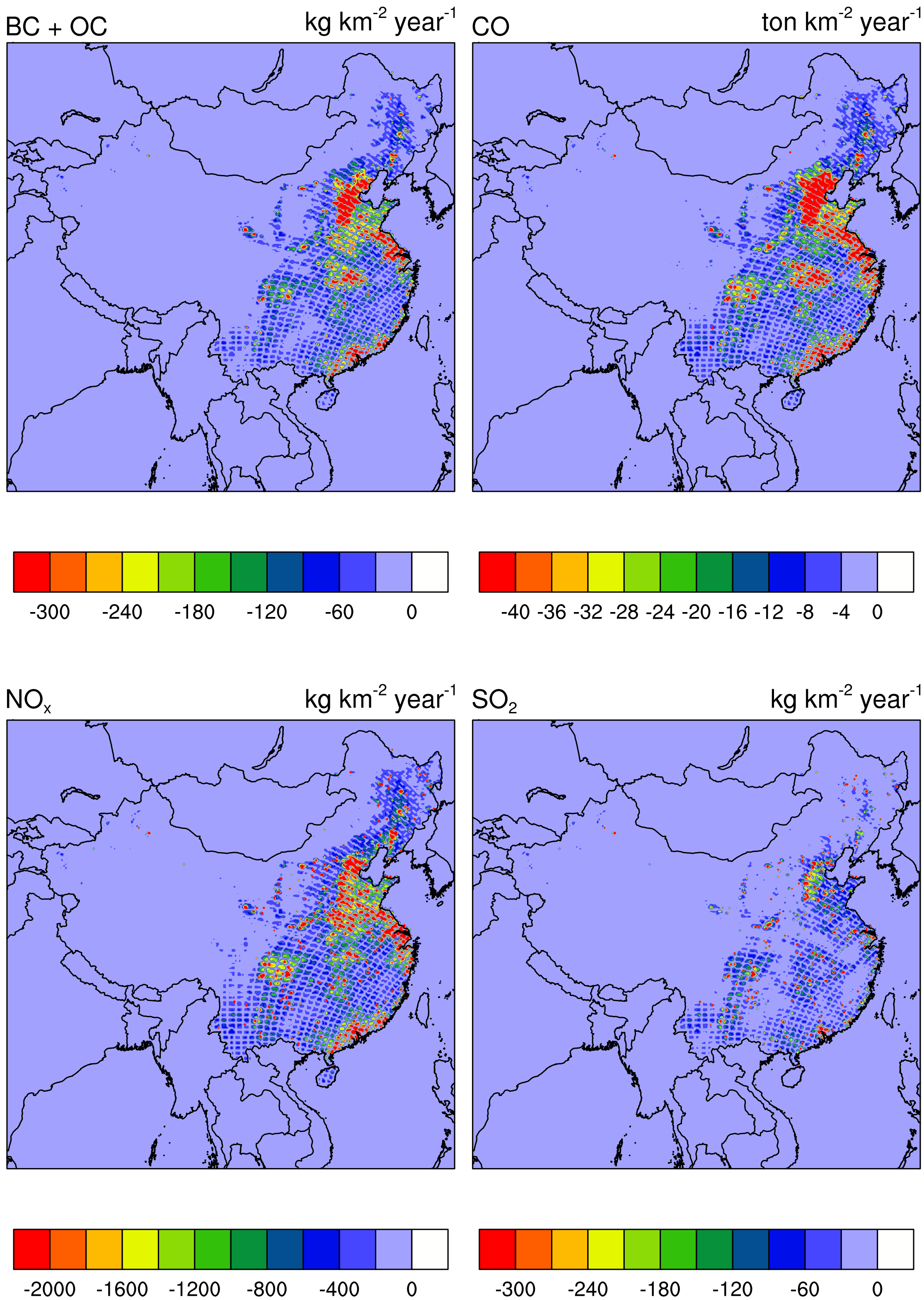


Figure 2-13: Spatial emissions differences of BC+OC, CO, NO_X, and SO_2 in 2020 between China 3/III (Scenario B) and no China 3/III (Scenario A′).

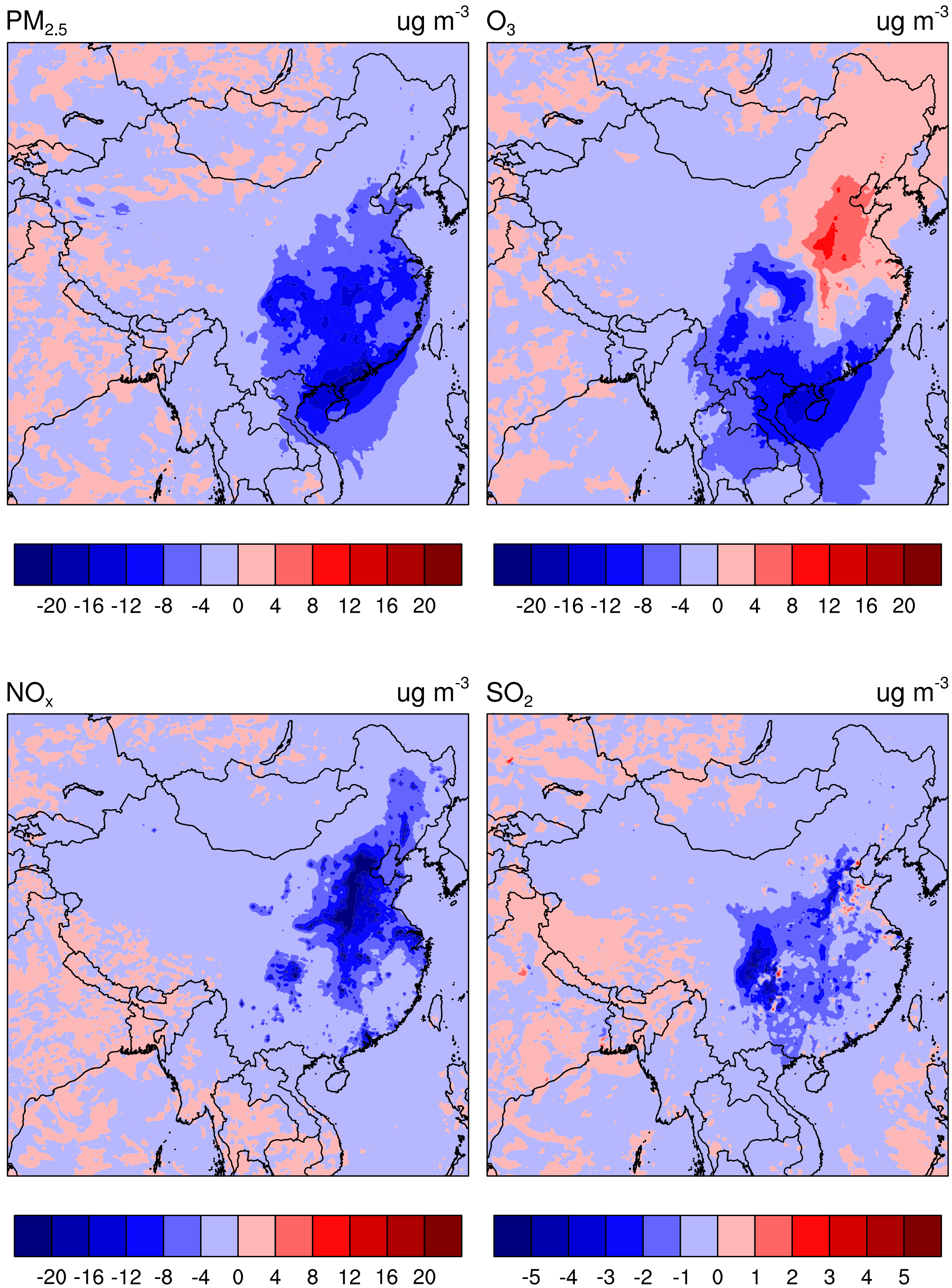


Figure 2-14: Concentration differences of $PM_{2.5}$, O_3, NO_X and SO_2 in 2020 between China 3/III (Scenario B) and no China 3/III (Scenario A′).

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Estimating transport demand by Chinese households and its relationship to urban characteristics

Abstract

In this essay, I provide new evidence about Chinese households' demand for transport. I estimate the recently-developed, flexible, Exact affine Stone index (EASI) demand system on urban household data from a survey with national coverage in three waves over the period 1995–2007. These data are augmented with indicators of urban form, local economic conditions, and transport system characteristics, capturing associations between these and households' budgets—particularly, the expenditure category of transportation and communication.

I find that transport expenditures are highly elastic with respect to total expenditure at the lowest incomes ($\hat{\epsilon}_x^{\mathtt{trn}} = 1.47$) and that this elasticity declines gradually with income but remains above 1. Transport spending rises monotonically from 1.6 % of budget at low incomes to 7.5 % at the highest. Income variation explains about one quarter of budget share variation across selected cities; the remainder except for error is associated with observable city-level measures and unobserved province- and year-specific attributes. Among the former, denser highway networks are linked to lower transport expenditure; while population density, city wealth, bus ridership, and taxi fleet size are associated with changes in overall household budgeting.

The findings provide a new perspective to complement demand elasticity literature that has focused on gasoline and vehicle-distance travelled (VDT), while the methods provide new means of connecting attributes of the built environment to welfare-consistent models of household expenditure.

3.1 Introduction

Rapid economic growth in the past three decades has increased incomes in China and transformed its cities. Households and individuals have responded by changing the way they travel, shifting from exclusively non-motorized transport, to taxis and public transit, privately-owned light-duty vehicles (LDVs), and long-distance rail and air. Along with demand, the external costs of transport have grown—including the health costs of air pollution and road injury; climate change effects of greenhouse gas emissions from fuel burning; and time lost due to congestion. Governments from the national to municipal level have responded to these issues with a diversity of policy measures, as discussed in Chapter 1. The design of these transport policies relies on knowledge about the relationship between transport demand, related economic drivers, and attributes of regions, cities, and transport systems that affect travel choices. This knowledge is used to anticipate counterfactual (no-policy) growth in transport activity, and to project or assess the impacts of policy. Several traditions of research have focused on different aspects of transport demand: household-level choice models based on transport-specific surveys and local data; regression models of national-level fuel demand elasticity based on aggregate statistics (both surveyed in detail in Section 3.2); and aggregate trajectories conditioned on local or international data (discussed previously, in Section 1.3).

In this chapter, I employ new methods and unconventional data to characterize Chinese households' transport behaviour by an important, yet less-often studied, measure: the share of expenditure devoted to transportation goods and services, alongside other categories of consumption within the overall household budget. Using urban household data from a large, social science survey with national coverage and three waves in 1995, 2002, and 2007, I estimate the recently-developed, Exact affine Stone index (EASI) demand system, a specification that allows the transport budget share to vary flexibly across the range of incomes. Estimation of EASI models yields Engel (budget share) curves, and estimates of the income elasticity of transport expenditure that also vary with income. As well—responding to prior research on relationships between travel and the built environment—I augment the household-level data with indicators of urban form, regional economies, and transport system characteristics, capturing the associations between these variables and households' budget shares, within and across categories of expenditure.

The chapter begins with a review (Section 3.2) of the literatures on estimation of household budget shares, demand elasticities, and the relationship between travel and the built environment. Section 3.3 focuses on systems of demand and applications to Chinese transport, and describes the gaps I address by adopting the flexible EASI form. In Section 3.4, I describe the three primary data sources used to produce a pooled data set of 18 624 observations of Chinese households. Section 3.5 gives the EASI model specifications used to investigate these data for effects suggested by the literature; as well as methods for model estimation and significance testing. In the remainder of the current chapter, I detail the findings from EASI models, their relationship to existing knowledge, and their implications in the Chinese context: Section 3.6 presents the relationship of transport demand with income through Engel curves and elasticities of expenditure and budget share; and associations with local conditions, through inspection of parameter estimates and a comparison of subsets of the data. In particular, I find that flexibility indeed allows demand systems to reflect transport budget shares that vary as higher order functions of income (or total expenditure); and that variation in income explains only a portion of overall changes in transport expenditure across China's provinces. Section 3.7 comments on implications and limitations of these findings, and concludes.

Separately, in the following Chapter 4, I evaluate the predictive power of the EASI demand system by exercising it against the simpler, commonly-used Almost Ideal demand system (AIDS) and testing sensitivity to the geographical coverage of data used for estimation. Measures of model fit are used to explore when and how modelers and practitioners could apply EASI in applications including projection and calibration of energy modeling frameworks; with other data; and also to suggest directions for future data collection.

3.2 Literature: characterizing demand in China and in cities

Chapter 1[1] gave a general introduction to models used to study transportation in China, with comment on levels of resolution; units and measures of analysis, and a critique of the use of aggregate functional relationships. Here, I survey two specific areas of transport research relevant to the quantities of focus in the current chapter, before turning in Section 3.3 to the details of demand systems: first, in Section 3.2.1, research that has sought to directly estimate elasticities of transport demands—for instance, gasoline or vehicle-distance travelled (VDT)—with respect to income or other quantities. This literature provides the few elasticity figures available for China, against which I compare results. Second, Section 3.2.2 covers research that has discussed the relationship between attributes of cities (the "built environment") and travel choices of individuals and households. Although not measured in terms of expenditure, these help form expectations (in Section 3.5.1) about the city-level observables included in the current work.

As discussed in Chapter 1, the CLIOS view of transport systems emphasizes that attributes of the system emerge from complex causal relationships. For the demand systems developed in this chapter, this suggests that measures of city-level characteristics may be endogenous with transport (or other) demands. In surveying the literature, I note throughout where and how prior research has flagged and/or responded to endogeneity in certain independent variables.

[1]in particular, Section 1.3 on page 28.

3.2.1 Elasticities of transport demands

The energy economics literature contains a plethora of estimates of the income elasticity of transport demands at the country level. Transport fuel (and specifically gasoline) demand has been the most commonly-studied concept, followed by measures of transport activity such as VDT—both usually expressed per capita. Some recent values are summarized in Table 3.1 on the following page.

China-specific estimates of income elasticity of gasoline demand have a broad range in the short run, from 0.160 to 1.77; longer-run estimates are closer to unity at 0.810 to 1.05. Dahl (2012), Goodwin et al. (2004), Labandeira et al. (2017), and Oum et al. (1992) provide surveys and meta-analyses, while P. J. Burke and Nishitateno (2013) perform analysis across a large numbers of countries, and Havranek and Kokes (2015) adjust Dahl's values for what they argue is a non-publication bias against low estimates.[2] Arzaghi and Squalli (2015) focus on fuel-subsidizing countries, including China, in a model with variables for weather, land area per capita, and percent urbanization. They argue that price controls in such countries ease concerns about the simultaneity of gasoline demand and price variables,[3] but do not address endogeneity or identification concerns for their other variables. P. J. Burke and Nishitateno address the endogeneity of price and quantity by instrumenting their models using, *inter alia*, fuel economy standards, vehicle import tariffs, and Kyoto Protocol membership. They note small changes in price elasticities, but no change in income elasticities of demand, of their IV models in comparison to base models.

For China in particular, C.-Y. C. Lin and J. Zeng (2013) produce linear regression estimates from province-level data, and find no significant estimate of the income elasticity of annual VDT per vehicle. They note that "it is quite difficult to determine

[2]Researchers finding insignificant or low-magnitude estimates of the elasticity are less likely to publish, resulting in a literature that contains only high and/or significant estimates.

[3]in turn, also critiquing Dahl, and P. J. Burke and Nishitateno for pooling subsidizing and non-subsidizing countries together.

Table 3.1: Literature estimates of elasticities of various measures of transport demand with respect to various variables. Discussed in Section 3.2.1 on page 85 (with respect to income), Section 3.3.2 on page 94 (AIDS methods), and Section 3.2.2 on the facing page (others).

Source	Elasticity of:	with respect to:	Value	Units & scope	Methods
		General, meta-analyses, or not including China			
Goodwin et al. (2004)	Total fuel consumption	Income	0.390[sr]	N = 175 meta-analysis; > 130 OECD, 0 China	Various regression & econometric; dynamic
	‘’	‘’	1.08[lr]		‘’; cross-section or
	‘’	‘’	0.490		time-series
Cervero and Murakami (2010)	VDT [km/(veh. year)]	Pop. density	−0.381	US; 370 cities; 2003 cross-section	Structural equation model (SEM)
	‘’	Road density	0.415		
	‘’	Hh. income[a]	0.209		
Dahl (2012)	Gasoline demand per capita	Income	0.230 to 2.06	124 countries	Lagged and non-lagged regression
Havranek and Kokes (2015)	Gasoline demand per capita	Income	0.100[sr]	124 countries	Dahl (2012) adjusted for non-publication bias
	‘’	‘’	0.230[lr]		
		China-specific, or including China			
McRae (1994)	Gasoline demand per capita	Income	0.220 to 1.77[sr]	11 Asian countries, national 1973–1987	Linear regression
C.-Y. C. Lin and J. Zeng (2013)	Gasoline demand per capita	Income	1.01 to 1.05	China provincial totals → national estimate	Lagged linear regression
	VDT [km/(veh. year)]	‘’	∼[ns]		
Cheung and Thomson (2004)	Gasoline demand per capita	Income	1.64[sr]	China, national/annual 1980–1999	Regression w/ cointegration
	‘’	‘’	0.970[lr]		
Arzaghi and Squalli (2015)	Gasoline demand per capita	Income	0.160[sr]	National, fuel-subsidizing countries incl. China	Lagged linear regression
	‘’	‘’	0.810[lr]		
H. Wang, P. Zhou, et al. (2012)	Transport	Expenditure	1.85	Provincial urban averages, 1994—2009	AIDS; national avg.
	‘’	‘’	1.20 to 4.20		‘’; provinces
Sun and Ouyang (2016)	Transport fuels	Expenditure	1.23	1032 households (CRECS), 2013	AIDS

[a] Median or average quantity across population.
[sr] Short-run.
[lr] Long-run.
[ns] Not estimated with significance.

appropriate instrumental variables for gasoline price," and elect to use diesel prices and international crude oil prices. Cheung and Thomson (2004), employing cointegration, note that "the dearth of data [for China] greatly constrained [their] choice of model." A recent study by Parker (2018), presenting non-parametric fuel demand estimates for non-OECD countries including China, uses variables for population density and urbanization. Parker explains that endogeneity in prices cannot be controlled for, because data for plausible instruments is not available. However—like Arzaghi and Squalli and the other studies mentioned—he does not discuss or treat endogeneity in these other variables.

3.2.2 Travel and the built environment

The urban studies literature on travel and the built environment contains extensive analysis of a large number of concepts and measures argued to be determinants of travel behaviour, itself measured in a fine-grained way. Ewing and Cervero (2010) provide one broad review and meta-analysis of this literature, focusing on more than 50 studies in the United States. Some of these studies report elasticity estimates; examples are included in Table 3.1. They describe the common categorization of local condition measures according a list of 'D's: density (of people or employment), diversity (i.e., mixed land use), design (e.g. of road networks), destination accessibility, and distance to transit. Each of these concepts has been operationalized and measured in multiple ways. For instance, destination accessibility can be measured as the number of jobs available within one mile of a respondents' home, or the distance to a downtown core; density can be measured using population, jobs, business, retail jobs, or other quantities.

A robust finding of this work has been that population density reduces measures (such as number of trips, trip share, distance traveled) of travel demand by motorized modes including LDV, while increasing non-motorized travel (walking and bicycling).

Gim (2012) meta-analyzes 35 studies and finds that the magnitude of the travel behaviour effect of density varies between studies of the United States and Europe, suggesting that such effects might also differ in other regions, such as China. Researchers often connect measured effects to policy recommendations: for instance, by linking density effects to energy use and emissions, Dulal et al. (2011) at the World Bank (WB) recommend increasing density as a means to mitigate transport greenhouse gas (GHG) emissions.

Many of the studies covered by such meta-analyses use household survey data. However, unlike the present chapter, the data are often from targeted, transport-specific surveys, either broad (covering multiple cities, regions, and/or transport modes) or more often narrow in focus—for instance, focusing only on one type of trip, such as non-work walking trips, or variation in a specific measure of destination accessibility. Dependent demand measures are often discrete, such as mode choice for a particular trip, or number of trips; so the resulting data are analysed using appropriate methods such as discrete choice models derived from random utility theory (Train 2009). Economic measures are commonly employed as independent variables (e.g. the costs of travel by certain modes), but expenditure is uncommon as a dependent variable. Most studies match measures of the built environment to individual observations based on respondent location, rather than relying on self-reported, or 'subjective' values furnished by respondents (Ewing and Cervero 2010).

Together with van de Coevering and Schwanen (2006), Ewing and Cervero critique one subset of the literature (e.g. Karathodorou et al. 2010; McIntosh et al. 2014; Newman and Kenworthy 1989) that uses aggregate dependent demand measures from databases of city-level data, since using these to draw conclusions about individual travellers in the same or other contexts constitutes an ecological fallacy. They recommend the use of either household data or adjustments for the "space-time context of cities" (van de Coevering and Schwanen 2006). However, across these and other studies, independent variables for local conditions are used at a broad range of resolutions from the neighbourhood to national level even when the dependent travel behaviors are measured for individuals or households.

Because causal relationships between concepts affect the validity of policy recommendations, this literature contains robust discussion of measures for treating endogeneity within the methods used, and for the demand and condition concepts examined. Cao et al. (2009) give a survey focused on responses to a single type of endogeneity: residential self-selection, which is the concept that households' choice of location is correlated with their travel choices. They identify at least nine methodological approaches used to control for this particular issue. In their discussion of instrumental variables responses (Cao et al. 2009, pp.367–377), they note that almost all attempts suffer from the difficulty of collecting suitable instruments; in most cases the instruments are weak[4] and thus errors in both directions are possible: significant effects may be erroneously found or estimated where none exist, or not detected where they do.

Another common methodological response noted by Cao et al. is the use of structural equation models (SEMs), in which a network of (possibly bi-directional) causal relationships is hypothesized, and estimation measures the strength of these links.[5] Cervero and Murakami (2010), for instance, use SEM to find a strong, negative elasticity of private-vehicle VDT per vehicle-year with respect to population density, and a positive elasticity with respect to road network density. However, they do not discuss, test or reject causation in the opposite direction.

Among the wide number of examples, P. Zhao et al. (2010, 2011) are an instance of recent work focused on China. The authors estimate discrete choice models of the

[4]Either only weakly correlated with the endogenous quantities they are instrumenting for, or not totally exogenous with the error term in model equations.

[5]Kline (2012, p.113) notes that SEMs were originally developed to be applied where some knowledge of causal direction exists *a priori*; these links are hypothesized to create a path diagram, and SEM can only disconfirm misidentified links. McIntosh et al. (2014), above, also use SEM.

commuting choice for individuals in households on the city fringe of Beijing, using both housing and location variables, and finding that compact urban form could reduce commute distances. They do not comment on or address potential endogeneity of their variables.

3.3 Demand system models

My objective in the present chapter is to estimate transport demand within a demand system that also captures other household consumption—one with rigorous, microeconomic foundations, which is flexible enough to capture the wide variation across households, provinces and time periods in a rapidly-evolving economy like China's. A related objective is to use household-level data that cover (or sample) the entire country, unlike the city- or district-specific models common in the literature. This section briefly introduces demand system concepts and the notion of flexibility (Section 3.3.1), touches on applications in China (Section 3.3.2), and describes the EASI demand system adopted here (Section 3.3.3).

This approach addresses limitations in the literatures identified above. Unlike previous transport demand elasticity estimates presented for China, the use of household data across a broad range of incomes and conditions increases statistical power and supports estimation of a flexible form in which data fully determine budget share parameters and elasticities. While the literature on travel and the built environment often measures nuanced concepts, data can be limited to particular cities and tailored survey questions; in contrast, my approach covers households sampled across the entire country and the variation in their local conditions, using common measures of expenditure that may be linked to aggregate economic models used for projection and policy analysis. Finally, unlike the stylized facts from aggregate functional relationships described in Section 1.3, demand systems impose overall income constraints, reflect individual households' observed, utility-maximizing choices in both transport and non-transport consumption, and do not rely on relationships derived from other countries.

3.3.1 Demand system concepts and flexibility

A demand system represents quantities demanded of two or more goods by an economic agent. In the context of partial equilibrium analysis, or computable general equilibrium (CGE) models,[6] such systems are derived from a production function such as the constant elasticity of substitution (CES) function, Equation (3.1).[7]

$$X = F\left(\sum_{j=1}^{n} a_j x_j^r\right)^{\frac{1}{r}} \tag{3.1}$$

where

$$\begin{aligned} X &= \text{aggregate consumption} \\ a_j &= \text{share of good } j \\ x_j &= \text{quantity of good } j \\ s = \frac{1}{1-r} &= \text{elasticity of substitution} \end{aligned}$$

By choosing X so as to maximize utility for a given total expenditure, y, the demand for each good, x_j, can then be expressed as a function $x_j \sim f(y, p_k \forall k)$ of y and the prices of each good, p_k. These demand functions together form the demand system. Restricted cases of the CES include the Cobb-Douglas function in which $s = 1$, and the Leontief function in which $s = 0$. The CES can be said to be more *flexible* than these, since the elasticity of substitution, s, may take on other values estimated from

[6]such as the Economic Projection and Policy Analysis (EPPA) (Chen et al. 2015) or China Regional Energy Model (C-REM) (Chapter 2).

[7]When applied to aggregate intermediate sectors, rather than final demand, firms choose a level of 'production' in order to maximize profit given market prices. When applied to final (consumer) demand, the thing 'produced' is utility.

data; however, the same value applies between any pair of goods or inputs when $n > 2$. 'Nested' or hierarchical CES structures were developed to permit a distinct s within each nest.

Another dimension of flexibility, and the one of interest here, is in income elasticity. Cobb-Douglas, Leontief, CES, and nested CES demands are homothetic, or homogenous of degree 1: when y is increased by some multiplicative factor, the demands for each good x_j increase by the same factor and remain in their original proportions. This represents the situation that the consumer does not alter her budget allocation in response to changes in income—only relative prices—and so the income elasticity of demand for each good is fixed to 1. Other functional forms were, in turn, developed to relax this restriction: for instance, the Stone-Geary demand functions have the form Equation (3.2).

$$x_j = \gamma_j + \frac{\beta_i}{p_i}\left(y - \sum_{k=1}^{n} \gamma_k p_k\right) \tag{3.2}$$

The parameters γ_j allow expression of a 'minimum' level of consumption of good j; by adjusting this parameter, the income elasticity of good j can be controlled at low incomes, though in the limit as y grows large the income elasticities tend to 1. More recently, researchers have developed demand systems including the AIDS (Deaton and Muellbauer 1980), "An implicit direct additive demand system" (AIDADS) (Rimmer and Powell 1996) and EASI (Lewbel and Pendakur 2009)—the last described in more detail below—that add additional flexibility.

Why might such flexibility be desirable? Figure 3-1 on the next page illustrates the situation through a variety of hypothetical Engel curves. In demand systems with constant income elasticity (red and blue in the diagram), an increase in household income brings a fixed response in terms of household budget share in a category of expenditure: none in the homothetic case (red), or a fixed increase or decrease (blue).

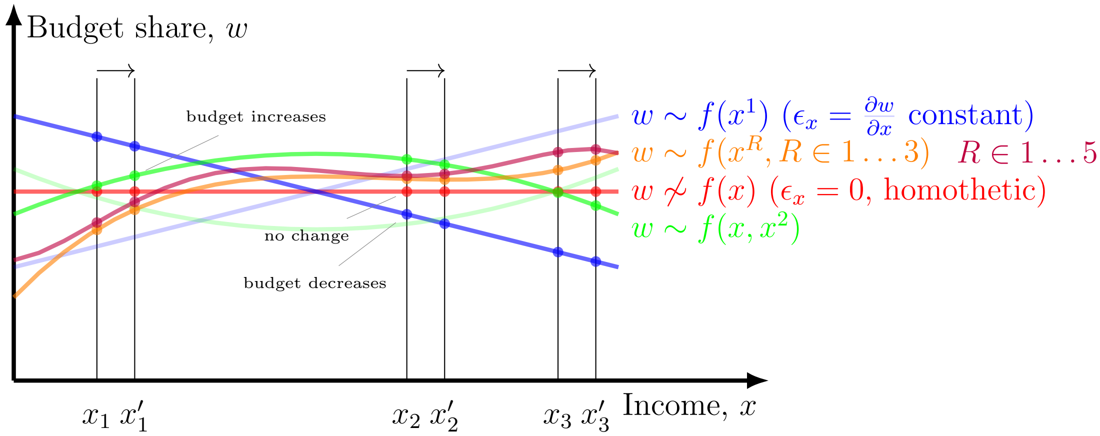


Figure 3-1: Functional forms for expenditure shares. Forms with higher-order functions of x capture distinct responses of households at different levels of income (x_1, x_2, x_3) to incremental shifts in income ($x_1 \rightarrow x_1'$, etc.) without imposing shape restrictions.

Quadratic forms (green) allow the response to vary, but with shape constraints: an elasticity > 1 (increasing budget share) at low incomes must be accompanied by a lower elasticity (descreasing budget share) at high incomes, or vice versa. Flexible forms (orange and purple), on the other hand, can allow data to more fully determine where elasticities change magnitude and/or sign; possibly multiple times across the range of incomes. Where data are plentiful, entirely non-parametric Engel curves can be described;[8] these forms, however, lack desirable qualities of rationality, and are more often used to explore parameterizations for later research (as in Parker 2018).

In the Chinese context of rapid growth, flexibility is a desirable feature because the available goods, their prices, incomes, and local conditions of households making budget decisions might well be expected to lead to very different consumption behaviour within samples spanning cities and years. Flexibility is also advantageous when dealing with aggregate categories, such as total transport expenditure in the data examined in this chapter: households in certain parts of the income distribution may respond to incremental changes of income by adjusting their spending on, for

[8]for instance, (Røed Larsen 2006) uses the voluminous data of the U.S. Consumer Expenditure Survey to draw non-parametric curves for subcategories of transport expenditure and individual demographic groups.

instance, long-distance airfare, while others might adjust private driving, or public transit use. As these goods and services have particular prices and households derive utility from them in distinct ways, total transport expenditure share will be affected differently by shifts at these different income levels.

3.3.2 Flexible demand systems: applications in China

Some of the newer, flexible demand systems have been applied in China. Research applying the AIDS to household data in China first focused on food, (e.g. Fan et al. 1995; Jiang and Davis 2007); these studies omit discussion of transport expenditure. More recently, Dai et al. (2012) estimate AIDS parameters using national aggregate data by income class, and project to consumption, energy use and GHG emissions in 2050, but do not report income elasticities; or incorporate variables to shift incomes. H. Wang, P. Zhou, et al. (2012) applied the linear approximation to the AIDS (LA-AIDS) using aggregate, province-level data in six categories of expenditure from 1994–2009, finding an income elasticity of transport expenditure of 1.85. Sun and Ouyang (2016) applied the LA-AIDS to consumption of three categories of energy goods using a household survey (China Residential Energy Consumption Survey (CRECS), $N = 1023$), finding an income elasticity of fuel expenditure of 1.23.[9] Caron et al. (2017) use a larger set of the CRECS data ($N = 4600$) to study energy goods consumption with a two-stage model that separately estimates the extensive and intensive margins of consumption; one of their categories is transport fuels, and they note that "our estimates imply relatively low income elasticities, except for gasoline and diesel." Finally, Z. Yang, Jia, et al. (2017) estimate an AIDS using aggregate, province-level gasoline and diesel demand, in which the subcategories are demand by particular types of vehicles, (from Huo, K. He, et al. 2011); thus they report, for instance, that diesel use by medium-duty trucks has an elasticity of 0.6 with respect

[9] these values are also included in Table 3.1.

to total expenditure in a province. Across these studies, measures of food and household energy demand have been the focus, with relatively little attention paid to total demand for transport or mobility, except per transport fuels. The present chapter contributes additional insights to stand alongside these, using new methods and data.

3.3.3 Exact affine Stone index (EASI) demands

I adopt the EASI demand system described by Lewbel and Pendakur (2009) and Pendakur (2009). This formulation is based on implicit Marshallian demands, and preserves rationality while allowing for unobserved preference heterogeneity and high rank (flexibility) in Engel curves. That is, the share of budget in each category may vary as a polynomial of utility, u, up to degree R, where $R \leq J$, the number of categories of expenditure. These features are useful for two reasons. First, a rational, or welfare-consistent set of demand functions can be used to estimate the welfare impact of constraints on consumption or changes in prices; this means that the models estimated here could be applied to analyze the aggregate cost of policies affecting transport expenditure. Second, while not entirely non-parametric, the high rank of the Engel curves permits variation in budget shares and income elasticities of demand for transport across the range of observed in incomes—allowing the data to inform as to the actual shape of these curves.

Let w^j be the budget shares for J consumption categories. The EASI budget shares (Equation (3.3)) are estimated using observations for individual households, i, as in Equation (3.4).

$$\frac{\partial \ln C(\mathbf{p}, u, \mathbf{z}, \epsilon)}{\partial p_j} = w^j(\mathbf{p}, u, \mathbf{z}) = m^j(u, \mathbf{z}) + \sum_{k \in J} \beta^j_{p,k}(\mathbf{z}) \ln p^k + \epsilon^j$$

$$m^j(u, \mathbf{z}) = \sum_{r=0}^{R} \beta^j_{u,r} u^r + \sum_{t \in T} \beta^j_{z,t} z_t \tag{3.3}$$

$$w^j_i = \sum_{r=0}^{R} \beta^j_{u,r} u^r_i + \sum_{t \in T} \beta^j_{z,t} z_{t,i} + \sum_{k \in J} \beta^j_{p,k} \ln p^k_i + e^j_i \tag{3.4}$$

The Hicksian[10] budget-share functions m^j depend on a vector of prices, $\mathbf{p}$, for the same categories; on polynomials up to order R of utility, u; and on a vector of demographic variables, $\mathbf{z}$, indexed by t. For ease of exposition I let the constant term be $u^r|_{r=0} = 1$ instead of a unit column in $\mathbf{z}$, and use labels instead of numerical indices for $j \in J$ and $t \in T$, with $|J|$ and $|T|$ being the sizes of these sets. The demands are *implicit* in u; Pendakur (2009) describes several methods to estimate the parameters $\beta^j_{p,k}\, \forall\, k$, $\beta^j_{u,r}\, \forall\, r$ and $\beta^j_{z,t}\, \forall\, t$, including the generalized method of moments, and iterated three-stage least squares regression (3SLS) with instrumental variables. In all approaches, an implicit estimate, y, of utility, u, is computed from total expenditure and a Stone price index (hence the name of the demand system). Using these estimators, the shares and parameters for one category's budget share are additively determined by the others; so I use the set $J' = J \setminus \{\texttt{other}\}$.

3.4 Data

Data for estimating the demand system are assembled from three sources. Each is described in turn: the variables w^j and total expenditure x from a national social science survey (Section 3.4.1); measures of city-level conditions $\mathbf{z}$ (Section 3.4.2) from a third-party database that collates official statistics; and prices $\mathbf{p}$ (Section 3.4.3)

[10] i.e., minimizing cost, $C(\cdot)$, for given utility, u.

using directly from the national statistical agency.

3.4.1 Household expenditure: the China Household Income Project

The China Household Income Project (CHIP) surveyed three populations of of households and individuals—rural, urban and rural-urban migrants—in five waves between 1988 and 2013 (S. Li et al. 2008; Luo et al. 2013). The survey used stratified sampling in repeated cross sections; the same provinces, prefectures and households were not re-visited in successive waves, so the observations do not constitute panel data. A slightly different survey instrument was administered to each of the target populations. For instance, the migrant survey sought to measure social connections formed in cities to which migrants had moved; the rural survey included sections to measure the impact of agricultural improvement policies.

The urban questionnaire for the 1995, 2002, and 2007 waves included questions on total household expenditure, spending in a varying number of subcategories, income, and some measures of assets. I adopt these as a primary data source, aggregating to the eight categories shown in Table 3.2 on the following page—including "transportation and communications"—and construct budget shares by taking the ratio with the total across categories.[11] I also construct, from the CHIP individual observations, a set of household-level descriptive variables listed in Table 3.3 on page 99. These describe either characteristics of the household—its size, measured in number of adults and children—or of the household head—including age, gender, marital status, and level of education.

Figure 3-2 on the following page gives the regional coverage of the CHIP urban data at the provincial level. Both wealthier eastern provinces and poorer central provinces were included, though the far western provinces (Xinjiang, Qinghai and

[11] For some observations, households' reported total expenditure is lower than the sum of reported expenditure in surveyed categories. In such cases, I discard the total, and use the sum instead.

Table 3.2: Expenditure categories

Category, j	Name
`clo`	Clothing
`dur`	Durables
`ed`	Education & recreation
`food`	Food
`hou`	Dwellings
`med`	Medical care
`other`	Other consumption
`trn`	Transportation & communication

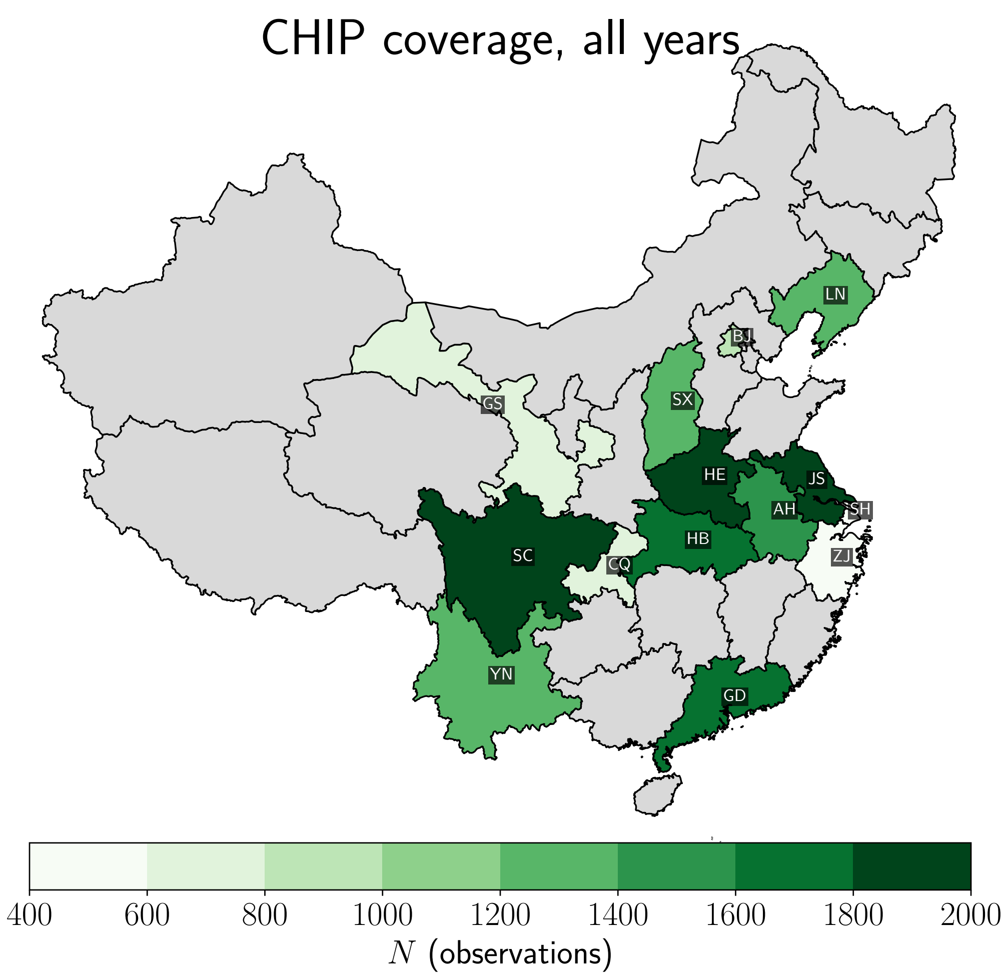


Figure 3-2: Number of observations by province in the entire CHIP data set, combined 1995, 2002, and 2007 waves.

Table 3.3: Household-level variables from CHIP data.

Variable	Description, CHIP response fields & condition 1995	2002	2007
`adult`[a,b]	Number of adult household members.		
	$\sum_i^N \begin{cases} 1 & \texttt{A5}_i > 16 \\ 0 & \text{o.w.} \end{cases}$	$\sum_i^N \begin{cases} 1 & \texttt{P106}_i > 16 \\ 0 & \text{o.w.} \end{cases}$	$\sum_i^N \begin{cases} 1 & 2007 - \texttt{a05\_1}_i > 16 \\ 0 & \text{o.w.} \end{cases}$
`child`	Number of child household members.		
		$N - \texttt{adult}$	
`age_years`	Age of household head, in years.		
	$\texttt{A5}_0$	$\texttt{P106}_0$	$\texttt{a05\_1}_0$
`age`	Household head is 40 y/o or older.		
		$\texttt{age} > 39$	
`educ`	Household head has completed more than middle school.		
	$\texttt{A11}_0 < 3$ [c]	$\texttt{P112}_0 > 4$ [d]	$\texttt{b02}_0 > 5$ [e]
`gender`	Household head is female.		
	$\texttt{A4}_0 = 1$	$\texttt{P105}_0 =$ 'female'	$\texttt{a04}_0 =$ 'female'
`single`	Household head is unmarried.		
	$\texttt{A7}_0 \in \{2, 3, 4, 5\}$	$\texttt{P109}_0 \in \{1, 3, 4, 5\}$	$\texttt{a07}_0 \in \{1, 4, 5\}$
`year_1995`[f]	Observation from the CHIP 1995 wave.		
	1	0	0
`year_2007`	Observation from the CHIP 2007 wave.		
	0	0	1

[a] Subscripts denote individual-level observations within a household, with 0 indicating the head of household. N is the number of individual-level observations in a given household.
[b] Response `a05_1` in CHIP 2007 gives birth year, rather than age.
[c] "middle level professional, technical or vocational school". Codes for this response are in descending, rather than ascending order.
[d] "Senior middle school (including professional middle school)".
[e] "senior middle school".
[f] We use these in place of the variable `year` = {0 if CHIP 1995, 1 if CHIP 2002, 2 if CHIP 2007} used by L. Li et al. (2015), which inappropriately implies linearity and equivalence of the 7- and 5-year periods between survey waves.

Table 3.4: CHIP sample counts, and rate in observations per million population by province and year.

	Observations, total			. . . per 10^6 pop.			
Code	1995	2002	2007	2000	2000	2010	Name
110000	492	484		39	34		北京市 Beijing
140000	649	640		21	19		山西省 Shanxi
210000	699	697		17	16		辽宁省 Liaoning
310000			499			24	上海市 Shanghai
320000	799	729	599	11	9	7	江苏省 Jiangsu
330000			586			11	浙江省 Zhejiang
340000	499	493	549	8	8	8	安徽省 Anhui
410000	599	680	641	6	7	6	河南省 Henan
420000	725	673	358	12	11	6	湖北省 Hubei
440000	540	544	685	7	6	7	广东省 Guangdong
500000		279	383		9	13	重庆市 Chongqing
510000	845	585	596	7	7	7	四川省 Sichuan
530000	647	636		16	14		云南省 Yunnan
620000	399	395		16	15		甘肃省 Gansu

Xizang (Tibet)) are omitted.[12] Total expenditure, x or `exp` (and thus $\log(\texttt{exp})$), increased across each wave, for both the entire data set and within individual provinces, as shown in Figure 3-3 on the facing page. The survey sampled different provinces and counties in each wave so that, for instance, observations are available from Beijing in the 1995 and 2002 waves, but not in 2007; or from Chongqing in the 2002 and 2007 waves, but not in 1995. Income distributions from provinces that appear in multiple waves overlap; e.g. there are observations of Guangdong households in 2002 that are wealthier than the poorest households sampled from the same province in 2007. Thus, there is no coverage gap in incomes, despite rapid economic growth in this period and the interval of up to seven years between CHIP waves. Table 3.4 gives the sampling rate in number of observations per million inhabitants at the provincial level for each year, varying between 7 and 39, with Beijing, Shanxi, and Shanghai most densely sampled.

Figure 3-4 on page 104 gives the distribution, across quantiles of income, for

[12] Appendix A on page 209 contains similar maps for each of the three CHIP waves used.

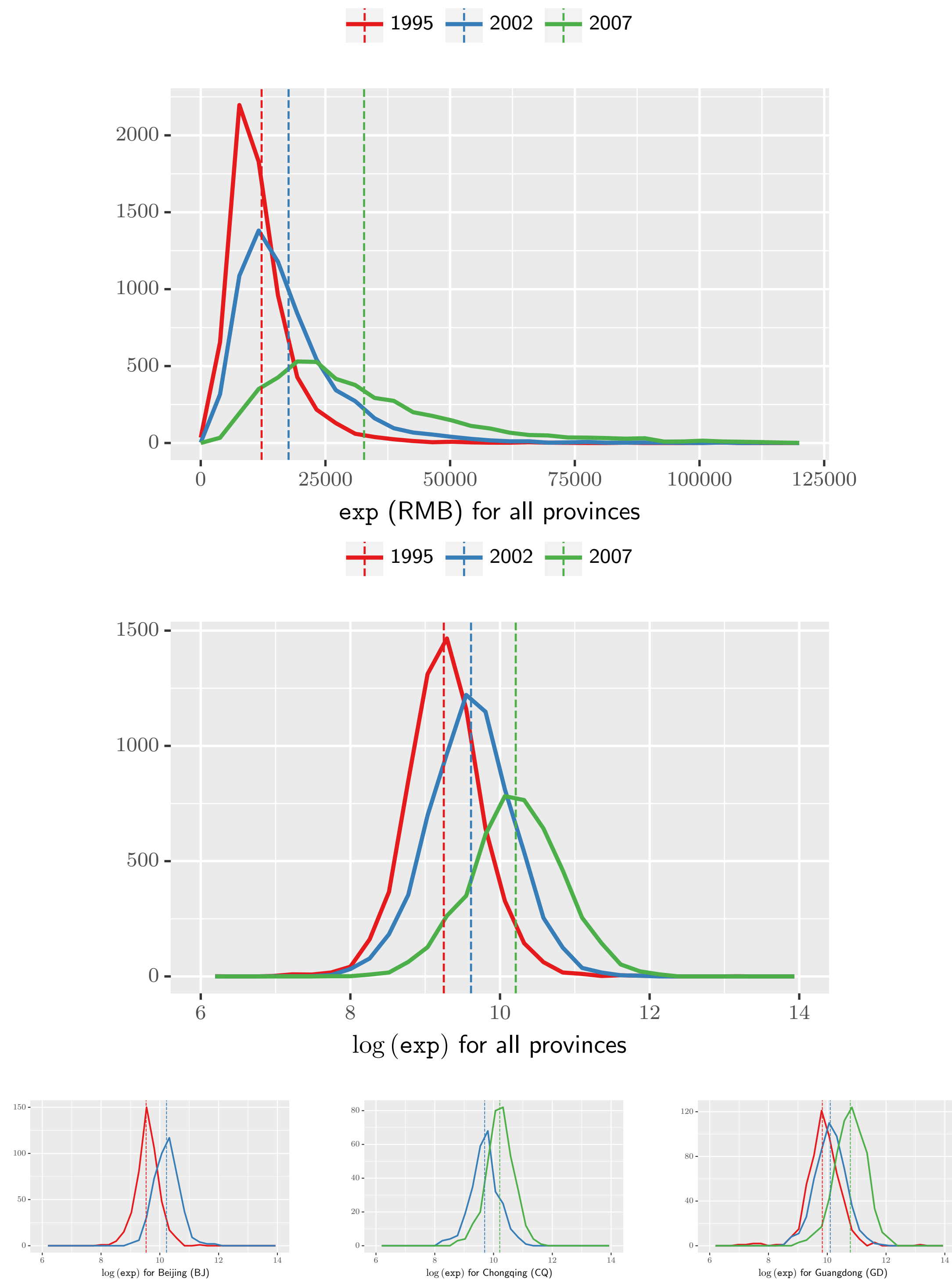


Figure 3-3: Distribution (solid line) and mean (dashed line) of total expenditure (top); and log (`exp`) (middle) for the entire data set; and of log (`exp`) for households in Beijing, Chongqing, and Guangdong (bottom row, left to right).

statistics on budget share of transport and communications, $w^{\texttt{trn}}$. Due to the large number of data points, moving statistics across 5 percent of the data are presented in this and subsequent plots. Several properties of the data are notable for estimation and interpretation of the results. First, since budget shares are fixed in the range [0, 1], the median (solid gray line) is closer to the first quartile than the third; and the mean (black) is higher than the median expenditure by 1 to 1.50 percentage points of expenditure. Mean budget share is drawn up by a number of high-side outliers; the treatment of these is addressed below.

Second, examining means from single CHIP waves reveals that reported transport budget shares are systematically lower in 1995 (red) than in 2002 (green) or 2007 (blue); however, 2007 shares are also lower than 2002 shares, at comparable levels of total expenditure. With reference to Figure 3-5 on page 105, which shows the share of observations in each wave across the overall distribution of the pooled data, the conditional mean expenditure in Figure 3-4 is seen to be closer to the 2002 and 2007 means at the highest expenditures, as there are comparatively few 1995 CHIP responses at this level. These differences have multiple potential sources. One is measurement: the CHIP survey instrument asked for a different set of subcategories in each year; and the way in which it was administered (for instance, any instructions given to respondents on what types of expenditure to include in the aggregate categories) may have led respondents to err differently in the different waves. Another is country-wide variation in conditions affecting transport consumption. For instance, in 1995, intercity rail service was generally at low speeds, and service was only gradually improved by policy campaigns during the 1995–2007 period (Hou and S.-M. Li 2011). Fixed, subsidized prices for transport fuels have also been adjusted periodically (B. Lin and Ouyang 2014). While the EASI demand models developed in this chapter allow inclusion of some measures that may reflect such changes, the budget share differences still suggest there may be influential, unobservable factors affecting differences in reported budget share, motivating the use of year fixed effects, as described in Section 3.5.2.

Returning to outliers in budget share, Table 3.5 on page 105 gives the fraction of these in each decile of total expenditure, by several methods. According to the standard boxplot metric, between 8 and 14 percent of observations in each decile are high-side outliers. Row (c) reports a notional threshold in absolute transport expenditure (rather than budget share): a household that purchased a vehicle in the year in which they were surveyed by CHIP may have (mis)reported by including this large lump sum as part of their average annual expenditure; however, gaps of this size are only visible in the highest decile.[13] In the lowest three deciles, a significant portion of households report no transport expenditure (row (d)); this may reflect all but exclusive use of non-motorized transport (foot and bicycle), which was widespread in 1995 and so cannot be identified as an error in data collection.

Because demand system estimates might be affected by erroneous data points not merely for $w^{\texttt{trn}}$, but for all w^j, I remove the most extreme high-side outliers in all expenditure categories by computing a quantile cutoff, q, according to Equation (3.5), and discarding observations where any budget share w^j is above the q-th quantile of shares in that category, using $k_q = 0.05$.

$$\min_q \operatorname{abs}\left(k_q - \frac{\left|i \in 1 \ldots N, j \in J : Q(w_j^j) > q\right|}{N}\right) \tag{3.5}$$

$$Q(w_i^j) \in (0,1] = \text{quantile of household } i\text{'s budget share in category } j$$

This has the effect of removing 5% of the data: as reported in row (a) of Table 3.5, from 2.90 to 6.30 percent in the lowest nine deciles, and 13.2 percent in the highest decile. This step is performed after augmenting the household-level observation with

[13] To illustrate the magnitudes in row (c), in a high income bracket in which 40 percent of households owned a vehicle that they replaced every 10 years, a randomly selected household might report such an expenditure with probability 0.04.

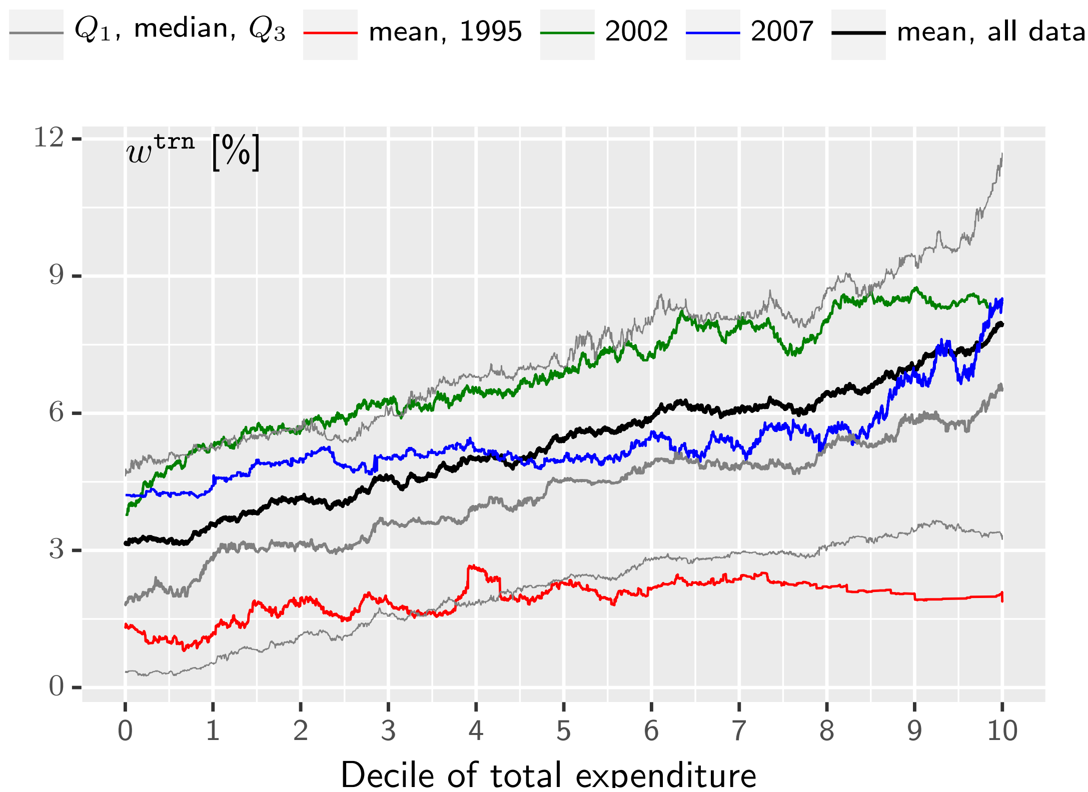


Figure 3-4: Moving average of $w^{\mathtt{trn}}$ for quantiles of `exp`, in black for all years, and in color for observations from the CHIP 1995, 2002, and 2007 waves. Median and first and third quartiles for all years in grey.

information from CEIC, below, and dropping observations for which the city-level variables are unavailable.

3.4.2 Urban and transport system characteristics: CEIC Data

In order to link the CHIP households' transport and other expenditures to the literature on travel and the built environment discussed in Section 3.2.2, I construct measures of urban and transport system characteristics from primary data of the National Bureau of Statistics of China (NBSC) and its provincial counterparts, as published in the annual China Statistical Yearbook, China Transportation & Communications Yearbook, and the analogous yearbooks of by each province's statistical bureaux. These data are publicly available, but not in a single, digital database, so I

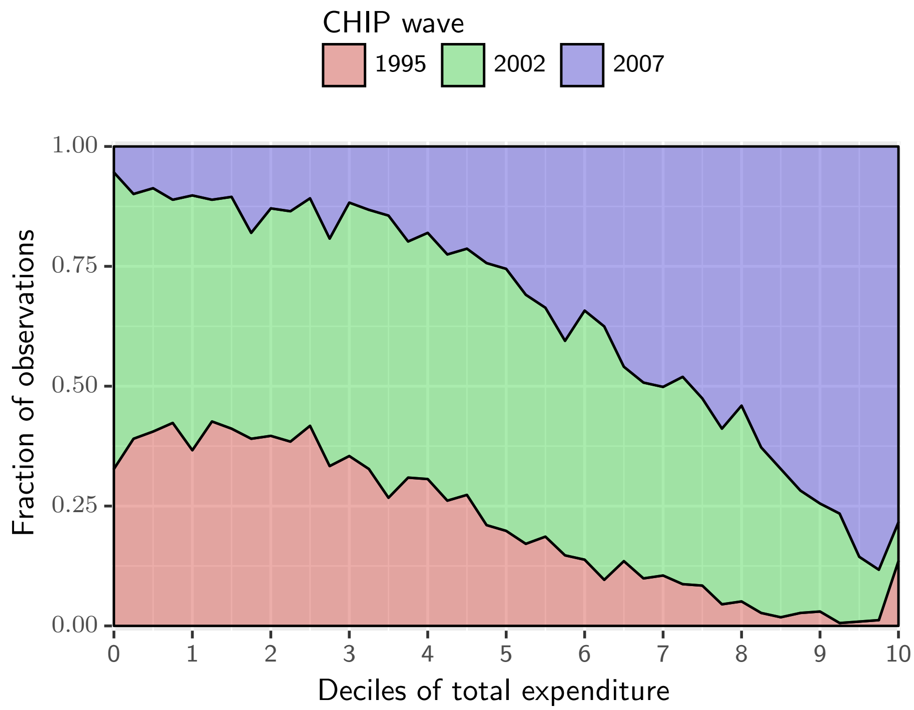


Figure 3-5: Fraction of observations by CHIP year across the range of expenditure in the pooled and cleaned data set. Roughly 35 percent of households in the lowest decile are from the 1995 CHIP wave; more than 75 percent of households in the highest decile are from the 2007 CHIP wave.

Table 3.5: Percentage of observations that are outliers in transport expenditure, by several metrics. (a) as described on page 103 for removing the 5% most extreme values across all categories; (b) standard boxplot metric of values greater than the median plus 3/2 the interquartile range; (c) absolute transport expenditure more than 10,000 RMB over the median (see Footnote 13 on page 103); (d) no transport expenditure reported.

Deciles of `exp`	1	2	3	4	5	6	7	8	9	10
(a) Censor 5% of obs., any j	5.8	4.0	2.9	3.0	3.3	3.2	4.0	4.5	6.3	13.2
(b) $w^{\texttt{trn}} > Q_2 + \frac{3}{2}(Q_3 - Q_1)$	13.6	9.8	8.6	9.8	9.9	8.1	8.6	9.6	8.7	10.0
(c) $\texttt{trn} > Q_2 + 10^4$ RMB	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.5	9.2
(d) No expenditure reported	28.0	16.8	10.7	7.1	4.6	3.3	2.0	1.2	0.9	0.8

use (CEIC Data 2016), a source that provides a compilation of the various yearbooks.

Variable selection was guided by two factors: relevance, and coverage. Among the many series available, I selected those most closely related to the measures identified by the literature as relevant to relationships between travel behaviour and the built environment, or to estimation of demand elasticities. The coverage of some of variables was limited to provincial totals or averages, or else did not extend far enough back in time to cover the 1995 CHIP wave. For instance, the total stock of commercial passenger vehicles (which might be used to compute a per capita measure) was only available for 32 province-level divisions. Such variables were omitted in favour of others that were available for provinces plus nearly 300 prefecture-level divisions including prefecture-level cities and districts in large cities, or for more than 2000 county-level divisions.

Table 3.6 on page 108 lists these variables, and gives the number of time-series available for each. The resulting set includes measures of the built environment in population density (`density`) and road network density (`hwy_density`);[14] of the local economy in gross domestic product (GDP) per capita (`gdp_cap`) and average wages (`wage_avg`); and the stocks of three kinds of transport vehicles (`stock_*_cap`). In Section 3.5.2, where I give model specifications, I discuss what the literature suggests about the relationship of the available variables to $w^{\mathtt{trn}}$. Some other attributes of the built environment discussed by the literature are not matched by measures available in the underlying official statistics thus the "China Premium Database" published by CEIC Data (CEIC) source: for instance, diversity of land use, and destination accessibility. These may be available from other public or private sources, or could be assembled manually, but I leave this for future research.

I assign these values to household-level CHIP observations by matching on the GB/T 2260 codes for administrative divisions, at the 6-digit county level where pos-

[14]Using the length of roads classified as expressways or Class I through VI highways, under the Chinese system (R. Zhou et al. 2016).

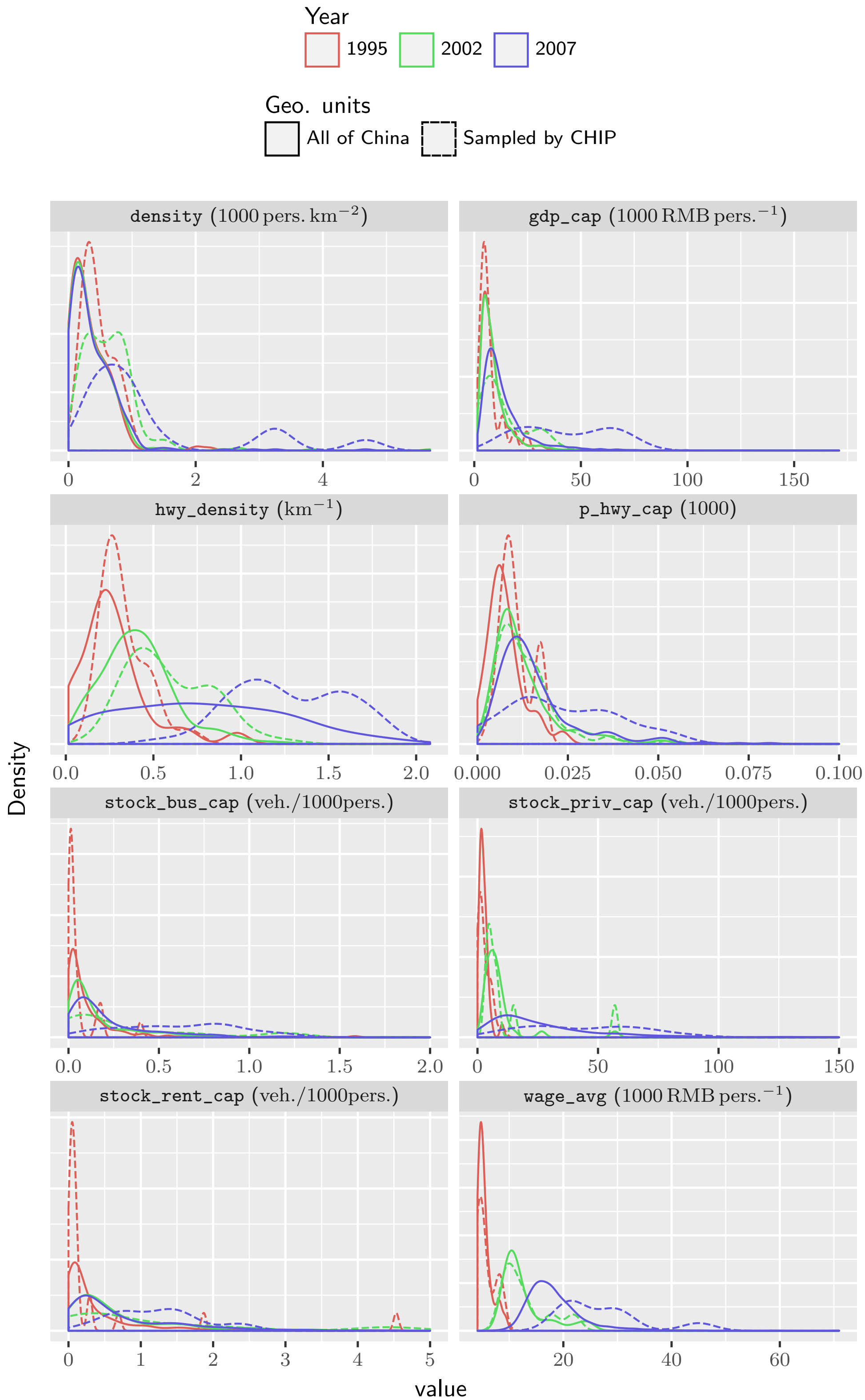


Figure 3-6: Distributions of urban variables: for all prefecture- or county-level divisions in China (solid lines) and for CHIP respondent households as matched by location (dashed lines).

Table 3.6: City-level variables. See also Figure 3-6 on page 107. '—' in the *Source* column indicates a derived variable

None	Description	Unit	Source	No. of series
Primary data variables				
wage_avg	Average Wage	10^3 RMB/pers.	NBS	1699
pop	Population	10^3 pers.	NBS	2382
hwy	Highway: Length of Highway	10^3 m	NBS, MoT	358
p_hwy	Highway: Passenger Traffic	10^6 pers.	NBS, MoT	314
stock_bus	No of Public Transit Vehicle: Bus and Trolley Bus	10^3 veh.	NBS	287
stock_rent	No of Rental Vehicle	10^3 veh.	NBS	287
gdp_cap	GDP: per Capita	10^3 RMB/pers.	NBS	2235
area	Land Area of Administrative Zone	10^6 m^2	NBS	359
gdp	GDP	10^6 RMB	NBS	2279
p_rail	Railway: Passenger Traffic	10^6 pers.	NBS	250
stock_priv	No of Motor Vehicle: Private Owned	10^3 veh.	NBS, MoT	325
Variables derived from primary data				
density	Population density	10^3 pers./km^2	—	359
gdp_cap_derived	GDP per capita, derived	10^3 RMB/pers.	—	2279
gdp_density	GDP density	RMB/m^2	—	330
hwy_density	Highway network density	1/km	—	358
p_hwy_cap	Passenger commercial road ridership per capita	10^3	—	314
p_rail_cap	Passenger rail ridership per capita	10^3	—	250
stock_bus_cap	Stock of buses & trolley bus vehicles per capita	10^{-3} veh./pers.	—	287
stock_priv_cap	Stock of private vehicles per capita	10^{-3} veh./pers.	—	325
stock_rent_cap	Stock of rental vehicles per capita	10^{-3} veh./pers.	—	287

Table 3.7: Price data series

Category, j	Code	NBSC description
`food`	A090202	Consumer Price Indices (preceding year=100), Food
`clo`	A09020G	[...] Clothing
`trn`	A090213	[...] Transportation and Communication
`hou`	A09021N	[...] Residence
`ed`	A09021D	[...] Entertainment and Education
`dur`	A09020M	[...] Durable Consumer Goods.
`med`	A09020R	[...] Health Care and Personal Articles
`other`	A090201	Consumer Price Index (preceding year=100)

sible, or at the 4-digit prefecture or 2-digit province level where more specific data is unavailable. Figure 3-6 on page 107 gives distributions of city-level measures across individual household observations in the resulting data set for the each of the three CHIP waves (dashed lines), in comparison to the distribution across the entire country (solid lines). Figure 3-15 (in Appendix 3.B, page 107) compares province-level means of these variables across the provinces covered (solid blue marks), and not covered (open black marks), also by year. Variation across and within years for these variables is sufficient for the parameter estimation that is the focus of the current chapter. Chapter 4 explores the implication mismatches between conditions in the provinces and cities used for model estimation, and those targeted for prediction and simulation.

3.4.3 Prices: the National Bureau of Statistics of China

Finally, I collect provincial- and city-level price indices in the same eight consumption categories as surveyed by CHIP, from the consumer price surveys of the National Bureau of Statistics of China (2008). Appendix B to the thesis describes the software

[15] Holz (2004, 2013), discussing the difficulty of compiling economic statistics in a transition economy, mentions that the sheer volume of prices to be collected mean that these data are a harder target for manipulation than data on output; consequently they are exempt from the suspicion sometimes cast on China's official GDP data, for which in any case Holz finds little confirmation using statistical tests.

used to retrieve these. The specific indices used for the consumption categories, including the codes assigned to the series by NBSC, are given in Table 3.7 on page 109. These top-level indices are assembled from sub-category indices; themselves based on frequent price surveys for specific goods at the third level of categorization.[15] Annual data by region and at the national level are published online. For demand system estimation, I transform prices by taking the cumulative product of the year-on-year indices, normalizing at the middle CHIP wave (2002), and taking the natural logarithm.

In addition to the eight top-level categories, I retain additional price series at the third level within the `trn` category, particularly `p_trn_fuel`, an index for the cost of road transportation fuels (diesel and gasoline). This variable is matched to household-level observations in the same manner as the CEIC variables described above.

The resulting, pooled data consist of 18 624 total observations covering cities in 14 provinces over three time periods, wherein each household's budget shares are augmented with same-year price information, household-level demographics, and measures of cities and local transport systems. I also add year- (refer back to Table 3.3 on page 99) and province-level indicators.

3.5 Models and methods

3.5.1 Explaining variation in transport expenditure

With the assembled data and demand systems just described, I estimate a variety of model specifications described below in Section 3.5.2, and apply statistical tests as described in Sections 3.5.3 and 3.5.4. Since the objective is to exploit the flexibility EASI to derive empirical facts from the CHIP data that describe the relationship of transport expenditure to incomes and local conditions, I first examine what expectations the literature suggests for the shape of transport Engel curves, and the relationship between $w^{\mathtt{trn}}$ and the urban variables listed in Table 3.6.

The parameter estimates for EASI models associate household-to-household variation in the dependent budget shares, w^j, with five kinds of independent variables in Equation (3.4):

1. the various powers of implicit utility, y^R (itself estimated implicitly from total expenditure, x, and prices),
2. household-level demographics from the CHIP (in $\mathbf{z}$),
3. city-level conditions from the CEIC source (in $\mathbf{z}$),
4. province- and year-level fixed effects (in $\mathbf{z}$),
5. prices from NBSC($\mathbf{p}$).

...with the remaining variation as an unexplained residual. From the perspective of the transport expenditure and demand elasticity literature, the results #1 give the shape of Engel curves and income-varying elasticities, as controlled for #2–5; while from the perspective of travel behavior, #3 give the association of local conditions with higher or lower spending, as controlled for the remainder. I discuss these in sequence below. A third matter of interest (addressed by the results in Section 3.6.3) is what portion of the budget share variation across households is explained by #1 vis-à-vis other sources: in other words, how strong is the influence of income versus observed and unobserved local, provincial, and year-specific conditions in households' budgeting for transport and generally?

Marchetti (1994) famously proposed "anthropological invariants in travel behaviour," in part based on fieldwork by Zahavi and Talvitie (1980) finding a travel money budget of 13 % across certain countries, 11 % in cities for vehicle owners, and 3 % to 4 % for non-car owners (Figure 3-7a on page 113). While the latter noted likely associations with income and car ownership, and possible association with urban structure, the former claimed that "personal travel is more under the control of basic instincts than economic drives"—which would imply in the current work that the variables describing urban structure will have no effect.

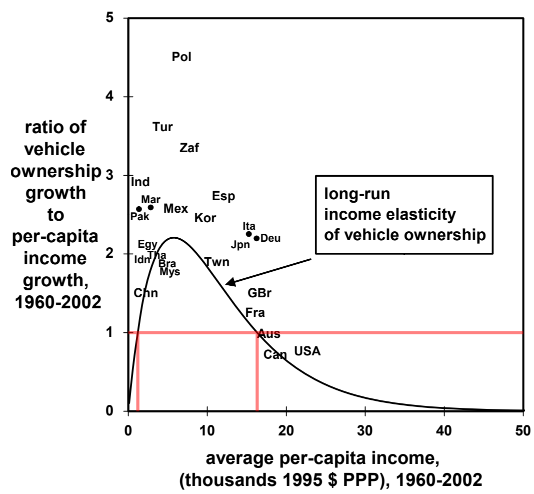
Though the demand elasticity literature, as summarized in Table 3.1, finds long-run values near 1 for gasoline in particular, the raw values in $w^{\texttt{trn}}$ already indicate a rising share that is below 3 % for the poorest households and near 8 % for the wealthiest, implying a value greater than 1. In contrast, Dargay et al. (2007) and others show that the income elasticity of vehicle ownership per capita rises and falls as a country passes the peak of its vehicle fleet growth, peaking well above 2 (Figure 3-7b on the next page). Ownership is related to vehicle purchase and fuel expenditures, though these are only a part of all transport expenditure, and only for some households; and as discussed on page 103, increases in ownership may be muted when measured by reported total transport expenditure, due to concurrent changes in spending on other transport goods and services when a vehicle is acquired, and the way in which the purchase is financed. However, given that China was approaching peak growth rates in vehicle sales by 2007, the year of the last CHIP wave in the data, an income elasticity greater than one is expected over at least part of the income range.

Regarding the urban variables, households in denser and larger cities may have access to more extensive public transit systems that lower the cost of travel, and thus budget shares. On the other hand, these may be wealthier households who spend less on basic needs such as food and clothing, and have a larger share of their income available for leisure travel on mode expensive modes. Thus the sign of coefficients on variables including `density` and `gdp_cap` could be positive, or negative. On the other hand, higher average wages may reflect a strong labor market in the city and thus, imperfectly, the accessibility of employment destinations. If high-pay jobs are available to households that are able to travel to them, higher `wage_avg` could be associated with increased transport expenditure or budget share.

As reported, Cervero and Murakami (2010) found a positive elasticity of VDT

NATIONWIDE vs. Total Household Expenditures, %			
US	1963–1975	13.2 ±0.38	
Canada	1963–1974	13.1 ±0.43	
UK	1972	11.7 ±0.38	
West Germany	1971–1974	11.3 ±0.54	
URBAN vs. Household Income, %			
		With Cars	*Carless*
Washington, DC	1968	11.0	4.20
Twin Cities	1970	10.1	3.40
Nuremberg Region	1975	11.8	3.50

(a) Figure 14 from Marchetti (1994). Original caption: "Rates of travel expenditure in various countries. Expenditure on travel appears to add up to quite a stable mean value of about 13% of *personal disposable income*. This budge [sic] is allocated between transport modes in a way that realizes maximum mean speed (i.e., territory). People who do not have a car use public services, which are usually underpriced, and in the available hour for travel appear unable to spend the whole budget."



(b) Figure 4 from Dargay et al. (2007), with unit elasticity noted. Original caption: "Historical Ratios of Vehicle Ownership Growth to Income Growth, by Levels of per-capita Income: 1960-2002"

Figure 3-7: Two examples from literature of regularities in transport demand.

per vehicle with respect to road network density. In the present models, the density of highway networks `hwy_density` could be correlated with either higher transport budget share, through the channel of a high road mode share, and consequently higher costs compared to public or non-motorized modes. Beyond the aggregate transport price index, higher fuel prices (`p_trn_fuel`) have been consistently related to a lower *quantity* of transport fuel purchased, in China (e.g. by C.-Y. C. Lin and J. Zeng 2013) and elsewhere. This knowledge does not lead to a straightforward expectation of the effect on total transport expenditure or budget share, $w^{\mathtt{trn}}$: households reducing quantity of gasoline purchased in response to price increases may still keep expenditure on gasoline constant; or they may shift consumption to other transport, or non-transport, goods and services.

Finally, variables more directly indicating mode share or which are aggregate measures of transport demand raise greater concerns about endogeneity, which may obscure the sign and magnitude of their associated parameters. Households in regions where the aggregate indicators suggest motorized road travel is heavier (higher `p_hwy_cap`, `stock_*_cap`) may spend more of their income to avail themselves of these modes and the associated mobility. On the other hand, the demand for taxi rides, for instance, certainly factors in the decision of drivers and firms to add additional for-hire vehicles (`stock_rent_cap`) and/or increase prices in order to maximize profits. In such cases, we might desire (analogous to P. J. Burke and Nishitateno (2013) and C.-Y. C. Lin and J. Zeng (2013), as discussed in Section 3.2.1) to instrument using a variable for some policy parameter that is correlated closely with the stock of for-hire vehicles, yet plausibly only loosely related to household-level budgets: for instance, the number of taxi medallions available in a particular city, or cost of a license, or any fixed fare. However, just as the CEIC source yields variables that do not precisely align with the literature on travel and the build environment, it also lacks data for these hypothetical instruments; and alternate data sources that systematize these policy variables across all the regions covered by the CHIP survey are not available. Consequently, any parameter estimates for these variables can only be interpreted with caution; I instead focus on where they are associated with changes in either transport budget share or overall budgets, by way of motivating future research.

3.5.2 Model specifications and estimation

I estimate EASI demand systems using the R software of Hoareau et al. (2012), with improvements and optimizations; data processing and model parameter and fit analysis are performed by new code. For details of the implementation, see Appendix B on page 229. Table 3.8 on the next page illustrates the variables included in the main models explored in this analysis; Table 3.16 on page 153 describes some additional models used for sensitivity testing. Source code for model specifications is given in Section A.2.1 on page 223. I use the following shorthand in model names to describe the regressors included:

"`yR`" Model contains R powers of implicit utility, y.

"`+hh`" Household level regressors for `age`, `educ`, `gender`, and `single`.

"`+city`" City-level regressors including `density`, `gdp_cap`, `hwy_density`, `p_hwy_cap`, `p_trn_fuel`, `stock_bus_cap`, `stock_priv_cap`, `stock_rent_cap`, and `wage_avg`.

"`-only-`*`var`*" Only `var` as a city-level regressor.

"`-many-dem`" A larger set of city-level regressors; see Table 3.16.

Because, for instance, `gdp_cap` ≈ `gdp/pop` and `density` is derived as `pop/area`, a regressor derived as `gdp_density` = `gdp/area` re-uses quantities already included in the regression and will absorb some of the effect attributed to the former two regressors. The variables in the "`+city`" set are selected to avoid such overlaps and partial multicollinearity. Unless otherwise noted, all models include both province- and year fixed effects. The former absorb unobserved, time-invariant, province-specific at-

Table 3.8: Specifications for main models. • indicates inclusion of a variable (columns) in a model. Other columns: number of observations, N; powers of y included in the estimation, R.

	N	R	density	gdp_cap	hwy_density	p_hwy_cap	p_trn_fuel	stock_bus_cap	stock_priv_cap	stock_rent_cap	wage_avg	age	educ	gender	single
y1	17689	1													
y1+city	13357	1	•	•	•	•	•	•	•	•	•				
y1+hh	17689	1										•	•	•	•
y1+hh+city	13357	1	•	•	•	•	•	•	•	•	•	•	•	•	•
y3	17689	3													
y3+city	13357	3	•	•	•	•	•	•	•	•	•				
y3+hh	17689	3										•	•	•	•
y3+hh+city	13357	3	•	•	•	•	•	•	•	•	•	•	•	•	•
y5	17689	5													
y5+city	13357	5	•	•	•	•	•	•	•	•	•				
y5+hh	17689	5										•	•	•	•
y5+hh+city	13357	5	•	•	•	•	•	•	•	•	•	•	•	•	•
y6	17689	6													
y6+city	13357	6	•	•	•	•	•	•	•	•	•				
y6+hh	17689	6										•	•	•	•
y6+hh+city	13357	6	•	•	•	•	•	•	•	•	•	•	•	•	•

tributes that affect w^j; the latter, year-specific attributes that affect all provinces simultaneously.

Before presenting the results of model estimation, the following two subsections give some additional methods necessary for interpretation.

3.5.3 Parametric correction for clustered regressors

Unlike previous applications of EASI (Hoareau et al. 2012; L. Li et al. 2015; Pendakur 2009), our method of adding city-level data to household-level observations introduces regressors that are clustered at the level of the prefecture in which households are located. Consequently, standard errors for $\hat{\beta}^j_{z,t}$—though not $\hat{\beta}^j_{p,k}$, $\hat{\beta}^j_{u,r}$, or t such as `age` that are household-specific—will be overstated. I therefore apply a parametric correction as described by Moulton (1986) (see also Angrist and Lavy 1999; Cameron and Miller 2015).

$$e^j_{ig} = \nu^j_g + \eta^j_{ig} \tag{3.6}$$

$$\rho^j = \frac{\sigma^2_{\nu^j}}{\sigma^2_{\eta^j} + \sigma^2_{\nu^j}} \tag{3.7}$$

$$\frac{V\left(\hat{\beta}^j_{z,t}\right)}{V_c\left(\hat{\beta}^j_{z,t}\right)} = 1 + \left[\frac{V\left(n_g\right)}{\bar{n}} + (\bar{n} - 1)\right]\rho_{z,t}\rho^j \tag{3.8}$$

Grouping households at the prefecture level, regression residuals e^j_{ig} for budget share in category j of observation i in group g are modeled as group- and individual-level components Equation (3.6) and their variances used to compute the intra-class correlation coefficient, ρ^j Equation (3.7). The ratio Equation (3.8) of the true ($V(\hat{\beta}^j_{z,t})$) and uncorrected ($V_c(\hat{\beta}^j_{z,t})$) parameter variance is computed using ρ; the variance ($V(n_g)$) and mean $\bar{n}$ of the group sizes; and the intra-class correlation coefficient of the regressors, $\rho_{z,t}$. In these models $\rho_{z,t} = 1$, since each household in the same prefecture receives the an identical value for the city-level regressors. To reiterate, the square root of this factor is applied to the standard errors of the parameter estimates for city-level variables, but *not* to the household-level variables or the powers of implicit utility, which are not clustered.

3.5.4 Tests for association with budget changes

The estimated demand system includes values and standard errors (some corrected as above) of $\hat{\beta}^j_{z,t}$ for *all* expenditure categories, j. Parameters for (notionally) transport-related variables such as `density` also appear in the equation for, e.g., $w^{\texttt{food}}$ ($\hat{\beta}^{\texttt{food}}_{z,\texttt{density}}$). It sometimes obtains that these parameters are estimated with significance according to a t-test, yet the corresponding parameter in the equation for $w^{\texttt{trn}}$ ($\hat{\beta}^{\texttt{trn}}_{z,\texttt{density}}$) is not. In this example—noting that `food` is the largest category of expenditure—the estimation result suggests that the variable in question *does* bear a relationship to households' consumption decisions; however, the effects is detectable only in certain categor(ies) of consumption, while indistinct in others, such as `trn`. However, since $\sum_j w^j = 1$, a variable that induces households to spend more (less) on, for instance, `food` also leaves them with less (more) for `trn` and other goods.

To formalize this observation, Section 3.6 includes tests of hypotheses such as Equation (3.9), which states that the variable t has no association with *any* budget share.

$$\mathcal{H}_0: \quad \beta^j_{z,t} = 0 \quad \forall \quad j \tag{3.9}$$

$$F = \frac{(SSE_R - SSE)(N-k)}{SSE \cdot q} \sim F_{(q,N-k)} \tag{3.10}$$

To perform these tests, the regressor of interest (e.g. $t = \texttt{density}$) is deleted and the resulting, restricted model is estimated. The F-statistic, Equation (3.10), is computed using the sum of squared errors from the restricted (SSE_R) and the base (SSE)

[16] for instance, model `y5+hh+city` has $R = 5$ powers of implicit utility, four household- and nine city-level variables in $\mathbf{z}$ ($|T| = 13$) and $J = 8$ categories of expenditure, so $k = 7(5 + 13 + 8) = 182$.

models, along with the number of observations (N); the number of parameters in the base model ($k = |J'|(R + |T| + |J|)$)[16] and the number of restrictions ($q = 7$ in these tests). Rejecting $\mathcal{H}_0$ allows us to say that the variable t *is* associated with—though not necessarily the cause of—changes in budget overall, even though the parameters $\hat{\beta}^{j}_{z,t}$ may be mixed in individual significance. These tests are also applied to the highest power of y in a given model; for instance, model `y2+hh+city` is model `y3+hh+city` with y^3 restricted; a significant value of F between these models will indicate that some or all of the Engel curves are at least cubic in implicit utilities.

3.6 Results and discussion

Estimation of the described models yields empirical information about household consumption that I discuss below in three categories, in each case contrasting results with the literature: first (Section 3.6.1) the fitted Engel (budget share) curves, showing that households' expenditure on transport rises from roughly 1.60 % to 7.50 % of budget across the range of incomes; second (Section 3.6.2), the elasticity of demand with respect to income, indicating that transport expenditure is strongly elastic ($\epsilon_x^{\mathtt{trn}} = 1.47$) at the lowest incomes and, while this elasticity declines, it remains significantly above 1 even at the highest incomes in the CHIP sample.

Finally, Section 3.6.3 on page 129 examines the relationships between city-level characteristics and transport expenditure through a comparison of two subsets of observations from different provinces and years, showing that the difference in these groups' mean $w^{\mathtt{trn}}$ is only partly explained by rising incomes, with city-level attributes, unobservables, and prices responsible for the remainder. The section concludes with an examination of the associations suggested by the underlying parameter estimates.

3.6.1 Engel curves for transport expenditure

I begin with the Engel curves for Chinese households' transport demand, which show a budget share for transport that is lower than suggested by the literature, but rises strongly across the range of incomes. Figure 3-8 on page 122 and Figure 3-9 on page 124 present the curves for, respectively, transport, and all categories, in the model `y3+hh`.[17] As with the raw CHIP data, moving statistics are shown including the first and third quartile, and the mean. By construction, estimated budget share for transportation and communications closely fits the conditional mean in the data, rising from 1.60 % at the lowest incomes, through 3 % at the second decile and 6 % at the eighth decile, to 7.50 % at the highest incomes. In contrast, Dai et al. (2012), applying the AIDS on national, aggregate data, estimate a transport budget share of 5 % in 1995 and 12.6 % in 2005 for urban households, and project a share of 26 % in 2050. While the shares in the EASI model are smaller, the increase in share of budget across the range of data used for estimation is much larger: a factor of 4.7, compared to 2.5 in Dai et al. (2012). Data from individual households, instead of aggregates at the mean of several large income categories, allows the models here to show a more complete picture of transport budget share variation from the lowest to highest incomes. Using the CRECS data, Caron et al. (2017) separate refined oil (gasoline and diesel) from other transport goods and services, showing respectively 2.10 and 2.60, and total 4.70 %[18] in 2007, nationwide across both urban and rural households. Using a flexible demand formulation to calibrate a CGE model, they project 2.90, 6.10 and a total of 9.00 % nationwide by 2030.

While the idea from Marchetti (cf. Figure 3-7a on page 113) of a fixed travel money budget (TMB) does not hold in my estimates, the 1968–1975 U.S. and European

[17] Because this presentation only shows variation across levels of $\log(\texttt{exp})$, the Engel curves for transport and other goods appear the same in other models, such as model `y5+hh+city` (Figure 3-14, in Appendix 3.B on page 154) that add city-level measures and additional powers of y.

[18] the summation here, not in the cited work, assumes that refined oil is solely used as a transport fuel.

figures from Zahavi and Talvitie (1980) for carless, urban households, at 3.40 % to 4.20 % of budget, fall within the range estimated here, while the figures of 10.1 % to 11.8 % for households with cars are beyond even the wealthiest households in the CHIP sample. Also outside of China, the results can be contrasted with the roughly contemporaneous, non-parametric results of Røed Larsen (2006) from U.S. households in 2000, with a mean budget share of 20.7 %, and decreasing budget shares across the range of incomes except for new vehicle purchases (families with ≥ 2 children) and air travel (single individuals in certain age ranges).

Turning to non-transport goods, Figure 3-9 shows only the conditional mean, for all j, illustrating how the parameter estimates reflect the differing shapes of the Engel curves: decreasing (e.g. `food`) or increasing (`ed`) in income; concave (`other`), convex (`hou`), or both in places (`dur`). A more detailed examination of the parameter estimates underlying the Engel curves reveals that the flexible demand system captures higher-order variation in transport budget shares across households. Table 3.9 on page 123 shows the full set of parameters for model `y3+hh+city`.[19] Categories of expenditure, j, are in columns, with parameters for `other` implied by the remainder and omitted. Parameter estimates with standard errors are in rows, with $\hat{\beta}^j_{u,r}$ (powers of implicit utility, including the intercept, $y^0 \sim r = 0$) followed by $\hat{\beta}^j_{z,t}$ (household and urban variables) and then $\hat{\beta}^j_{p,k}$ (prices). Estimates are scaled to percentage points of budget share; for instance, the constant term for `trn` of 7.06 (standard error 2.45) indicates that $w^{\texttt{trn}}$ is 7.06 % at zero income and zero values of the other variables; the estimate is significant at the 1% level.

Parameters estimated as significantly different from zero in one budget share equation are not necessarily so in others; this is examined in more detail in Section 3.6.4, but note in particular that the parameter for the third power of implicit utility, $\hat{\beta}^j_{u,3}$, is significantly different from zero in all budget shares except durable goods (`dur`). For

[19]Similar tables are shown in Appendix 3.A for model `y5-many-dem` (Table 3.13), model `y3+hh` (Table 3.12), and model `y5+hh+city` (Table 3.14).

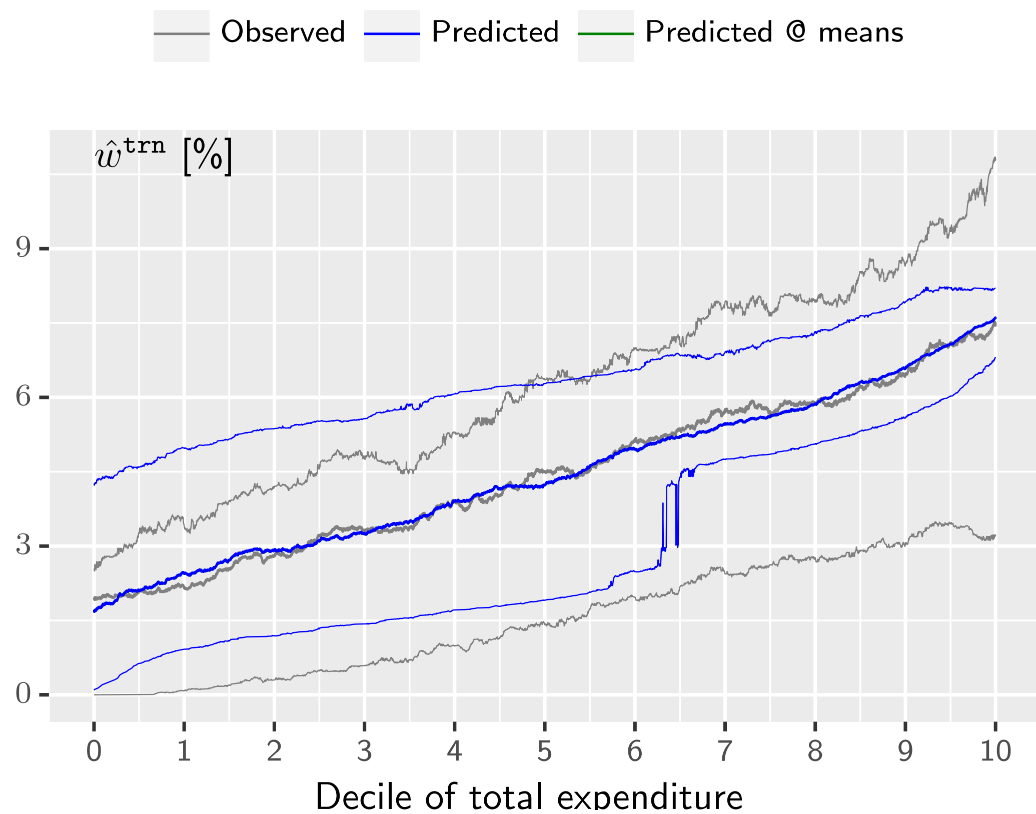


Figure 3-8: Fitted `trn` budget share (blue) for model `y3+hh`: moving average (heavy lines) and first and third quartiles (thin lines), versus same quantities in raw data (grey, as in Figure 3-4 on page 104).

Table 3.9: Estimated Engel curve parameters for model `y3+hh+city`. * = estimate significant at the 10% level, ** = 5%, *** = 1%.

	`food`	`clo`	`trn`	`hou`	`ed`	`dur`	`med`
			Powers of (implicit) utility				
$\hat{\beta}^j_{u,0}$	−16.738**	−8.979***	7.061***	−0.655	8.499*	12.921***	63.820***
	(7.564)	(3.337)	(2.446)	(3.178)	(4.540)	(2.888)	(3.645)
$\hat{\beta}^j_{u,1}$	33.043***	−2.543**	−4.230***	14.599***	−10.374***	−3.638***	−22.163***
	(3.049)	(1.201)	(0.926)	(1.316)	(1.902)	(1.138)	(1.451)
$\hat{\beta}^j_{u,2}$	−4.198***	0.474**	0.650***	−2.482***	1.469***	0.381**	2.649***
	(0.476)	(0.188)	(0.145)	(0.206)	(0.297)	(0.178)	(0.227)
$\hat{\beta}^j_{u,3}$	0.139***	−0.019**	−0.025***	0.116***	−0.052***	−0.005	−0.102***
	(0.022)	(0.009)	(0.007)	(0.010)	(0.014)	(0.008)	(0.011)
			Household-level variables				
$\hat{\beta}^j_{z,\texttt{age}}$	1.815***	−1.903***	−0.427***	−0.532***	−0.467***	0.006	1.083***
	(0.269)	(0.106)	(0.082)	(0.116)	(0.168)	(0.100)	(0.128)
$\hat{\beta}^j_{z,\texttt{educ}}$	−4.623***	0.393***	0.392***	0.061	−0.334**	−0.035	−0.435***
	(0.252)	(0.099)	(0.077)	(0.109)	(0.158)	(0.094)	(0.120)
$\hat{\beta}^j_{z,\texttt{gender}}$	−1.193***	0.544***	0.030	−0.224**	0.401**	−0.062	−0.370***
	(0.251)	(0.099)	(0.076)	(0.109)	(0.157)	(0.094)	(0.120)
$\hat{\beta}^j_{z,\texttt{single}}$	0.662	−0.321*	0.270*	0.488**	−0.302	0.441**	0.843***
	(0.471)	(0.185)	(0.143)	(0.203)	(0.294)	(0.176)	(0.224)
			City-level variables				
$\hat{\beta}^j_{z,\texttt{density}}$	2.798***	0.804**	0.164	−0.521	−0.076	0.068	0.162
	(0.965)	(0.324)	(0.235)	(0.395)	(0.431)	(0.255)	(0.307)
$\hat{\beta}^j_{z,\texttt{gdp\_cap}}$	−0.123*	−0.034	−0.010	0.039	0.021	−0.025	−0.015
	(0.065)	(0.022)	(0.016)	(0.028)	(0.031)	(0.017)	(0.021)
$\hat{\beta}^j_{z,\texttt{hwy\_density}}$	1.802	−0.945	−1.263*	−2.317**	−0.694	0.317	0.035
	(2.807)	(0.981)	(0.699)	(1.120)	(1.200)	(0.730)	(0.888)
$\hat{\beta}^j_{z,\texttt{p\_hwy\_cap}}$	153.717**	17.419	6.389	−2.561	−30.701	11.495	−0.431
	(64.227)	(21.931)	(15.508)	(26.700)	(28.951)	(16.567)	(20.729)
$\hat{\beta}^j_{z,\texttt{p\_trn\_fuel}}$	15.206	−24.207	−1.213	14.794	−11.287	9.318	6.852
	(30.968)	(15.534)	(9.237)	(12.149)	(11.511)	(8.761)	(11.167)
$\hat{\beta}^j_{z,\texttt{stock\_bus\_cap}}$	−0.681	−2.583**	0.119	2.634*	−2.616*	−0.082	0.055
	(3.241)	(1.099)	(0.800)	(1.384)	(1.496)	(0.872)	(1.042)
$\hat{\beta}^j_{z,\texttt{stock\_priv\_cap}}$	−0.071	0.023	0.008	0.005	0.003	−0.005	0.029*
	(0.053)	(0.019)	(0.013)	(0.022)	(0.024)	(0.014)	(0.017)
$\hat{\beta}^j_{z,\texttt{stock\_rent\_cap}}$	1.858	0.784*	0.310	−0.280	1.004*	−0.132	0.152
	(1.265)	(0.423)	(0.307)	(0.536)	(0.588)	(0.335)	(0.405)
$\hat{\beta}^j_{z,\texttt{wage\_avg}}$	−0.270	0.650***	0.121	−0.384*	0.596**	−0.197	−0.086
	(0.581)	(0.247)	(0.161)	(0.232)	(0.246)	(0.159)	(0.192)
			Own- and cross-price elasticities				
$\hat{\beta}^j_{p,\texttt{food}}$	71.211***	17.810***	28.643***	19.908***	27.514***	−6.056	50.905***
	(13.873)	(5.096)	(4.326)	(3.591)	(4.033)	(4.275)	(6.204)
$\hat{\beta}^j_{p,\texttt{clo}}$	17.810***	−7.440*	5.855***	7.278***	−4.380**	9.547***	6.955**
	(5.096)	(3.930)	(2.230)	(1.714)	(1.965)	(2.288)	(2.779)
$\hat{\beta}^j_{p,\texttt{trn}}$	28.643***	5.855***	11.863***	3.382**	2.744*	−5.246***	11.466***
	(4.326)	(2.230)	(2.538)	(1.342)	(1.461)	(1.901)	(2.215)
$\hat{\beta}^j_{p,\texttt{hou}}$	19.908***	7.278***	3.382**	−1.281	10.716***	−0.380	5.421***
	(3.591)	(1.714)	(1.342)	(1.615)	(1.423)	(1.492)	(1.949)
$\hat{\beta}^j_{p,\texttt{ed}}$	27.514***	−4.380**	2.744*	10.716***	−1.774	−4.799***	9.115***
	(4.033)	(1.965)	(1.461)	(1.423)	(2.635)	(1.648)	(1.923)
$\hat{\beta}^j_{p,\texttt{dur}}$	−6.056	9.547***	−5.246***	−0.380	−4.799***	−1.713	−14.194***
	(4.275)	(2.288)	(1.901)	(1.492)	(1.648)	(2.660)	(2.139)
$\hat{\beta}^j_{p,\texttt{med}}$	50.905***	6.955**	11.466***	5.421***	9.115***	−14.194***	13.102***
	(6.204)	(2.779)	(2.215)	(1.949)	(1.923)	(2.139)	(3.697)

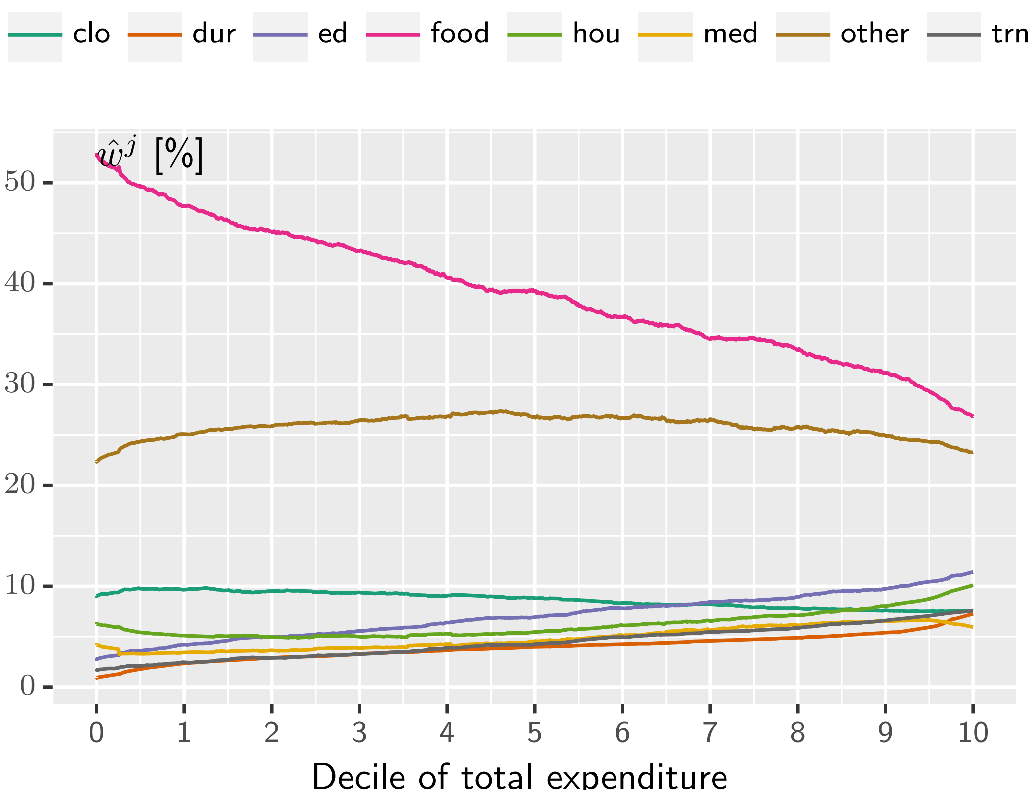


Figure 3-9: Fitted budget shares, model `y3+hh`.

Table 3.10: Comparison of several models. Significance indicators for t-tests are for parameters in the budget share equation for `trn`; indicators for F-tests are across all budget shares, per Section 3.5.4.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Model name	`y3`	`y3+hh`	`y3+hh+city`	`y3+hh+city`	`y3+hh+city`	`y3+hh+city`	`y4+hh+city`	`y5+hh+city`	`y6+hh+city`	`y6+city`
Observations, N	17689	17689	13357	13357	13357	13357	13357	13357	13357	13357
Household vars	–	4	4	4	4	4	4	4	4	–
Year fixed effects	–	–	–	•	–	•	•	•	•	–
Province fixed effects	–	–	–	–	•	•	•	•	•	–
Powers of (implicit) utility (t, or t / F, tests)										
R	3	3	3	3	3	3	4	5	6	6
0	***	***	***	***		***				
1	***	***	***	***	***	***				
2	***	***	***	***	***	***	**			
3	**	**	***	***	***	***/***	**			
4	–	–	–	–	–	–	***/***			
5	–	–	–	–	–	–	–	/		
6	–	–	–	–	–	–	–	–	/	/
City-level variables (t / F tests)										
Number, including FEs	–	–	9	11	22	24	24	24	24	9
`density`			/***	* /***	/***	/***	/***	/***	/***	/***
`gdp_cap`			/***	/***	/***	/ **	/ **	/ **	/ **	/***
`hwy_density`			/***	* /	/	* /	* /	* /	* /	/***
`p_hwy_cap`			/***	/***	/***	/***	/***	/***	/***	/***
`p_trn_fuel`			***/***	/***	***/***	/	/	/	/	***/***
`stock_bus_cap`			/***	/***	/	/	/	/	/	/***
`stock_priv_cap`			/ **	/***	/	/	/	/	/	/***
`stock_rent_cap`			/***	/***	/	/ *	/ *	/ *	/ *	/ **
`wage_avg`			***/***	***/***	***/***	/	/	/	/	***/***

are dropped (column (10)), significance does not return for any powers of y.

3.6.2 Income elasticities of budget share and expenditure

I find that, as households' income increases, their response is uniformly to increase not only their absolute (RMB) expenditure on transport, but the share of budget for this category. While demand is strongly elastic, the range of elasticities across the range of incomes is not as large as previous research, based on aggregate data, has suggested. Figure 3-10 on page 130 shows the elasticities of transport budget share, $w^{\mathtt{trn}}$ (top) and of `trn` expenditure (bottom) in the model `y5+hh+city`, giving the median (black) and first and third quartiles (thin grey lines).[20] Transport and communication expenditures are elastic with respect to income throughout the range of incomes in the CHIP survey. This elasticity is highest at the 7th percentile of income, at 1.47; reaches 1.30 for median income, and then falls to 1.06, just above unity, at the highest incomes. In contrast, the transport budget share elasticity reaches a maximum of 0.0142 at the 22nd percentile of income, is 0.0138 at median income, and 0.004 68 at the highest incomes. The high share elasticity, remaining roughly constant up to the eighth decile, shows that households across this range respond to rising income with the same increase in the fraction of their spending that is devoted to transport.

Returning to Table 3.1, the estimates of H. Wang, P. Zhou, et al. (2012) offer the closest analogue to the quantities modeled here. In order to investigate rebound effects of energy efficiency improvements, they used the linear approximation to the AIDS with aggregate, provincial data to find a nation-wide average transportation expenditure elasticity of 1.85. Their per-province estimates span a large range, from 1.2

[20]In demand system models where the dependent variables are budget shares w^j, income elasticities of demand may be expressed in one of two ways: either the elasticity of a budget share (e.g. $w^{\mathtt{trn}}$) *or* elasticity of total expenditure in a category (e.g. `trn`), with respect to income (Hoareau et al. 2012). Lewbel and Pendakur (2009) refer to these as 'semi-elasticities' and 'elasticities' respectively; here, I use the terms 'share-' and 'expenditure elasticity'. Across a population with homoth-

(Guizhou) to 4.2 (Yunnan), and do not clearly correspond to the provincial average incomes. For instance, their estimates for the wealthy, direct-controlled municipalities of Beijing, Shanghai, and Tianjin are 2.4, 3.0, and 1.7, respectively—all values far above the range of the EASI estimates. The differences point to two advantages of the present approach. First, estimates on provincial aggregate data can include only statistics, and not individual values, for local condition and household demographic variables (and in the application of H. Wang, P. Zhou, et al. do not incorporate any such controls). As such, they may associate a larger share of changes in transport expenditure with rising incomes. This effect is readily reproduced: an EASI model (`y5`) that deletes household-level demographics, city-level measures, and both province and year fixed effects gives a higher peak elasticity estimate of 1.85. Second, aggregate data obscure the distributions of income and expenditure in the total and within the transport category. Shifts in the shape of this distribution (cf. Figure 3-3 on page 101) could lead to changes in expenditure for households in the tails, while central measures, such as the mean, of income or wealth change only slightly. Estimates based on disaggregate data instead reflect the observed behaviour of these households individually.

The other China-specific elasticity estimates from Table 3.1 are not directly comparable to the current results, as they measure different dependent quantities, such as gasoline demand. Sun and Ouyang (2016) give an elasticity of expenditure on transport fuels of 1.23 based on the CRECS; this is not inconsistent with the present results, as fuels are but one good among other goods and services in the transport basket. Supposing this value held across the range of incomes, then it would imply a higher expenditure elasticity for non-gasoline transport goods and services where the EASI $\hat{\epsilon}^j_{x,i}$ is lower than this value, and vice versa. While C.-Y. C. Lin and J. Zeng

etic preferences (cf. Figure 3-1 on page 93), the *expenditure* elasticity is constant at unity: category expenditure grows in direct proportion with total income. It is equivalent to say that the *share* elasticity is constant at zero: the share of budget in the category does not change with income.

(2013) find no significant elasticity of a demand measure (VDT per vehicle per year) with respect to income, the results here show that demand measured as expenditure is significantly elastic.

Finally, for households in the 2000 U.S. Consumer Expenditure Survey, Røed Larsen (2006), gives an elasticity of 0.74 for overall transport expenditure, and also finds inelastic behavior for "necessary goods of transportation," such as local public transportation (including mass transit), and vehicle insurance and maintenance. On the other hand, he finds income-elastic ($\epsilon > 1$) demand for purchases and leases of new vehicles, as well as expenditures on leisure travel, including intercity trips by air and rail. While the CHIP survey did not record these components, low-income households in China are less likely to own vehicles, take leisure trips, and air transport, and more likely to be reliant on public transport for mobility; yet I find the demand of these lower-income households to be *more* elastic. Even if this difference is attributable to the gap between 2000 U.S. income distribution and that of the CHIP respondents—Røed Larsen uses a lower bound of 2×10^4 USD for Engel curves, which is above the 90th percentile of CHIP incomes—it highlights the importance of country-specific demand estimates.

There are two further points to be made about the EASI elasticity findings. The conditional distribution of the expenditure elasticity, $\hat{\epsilon}^{j}_{x,i}$, in Figure 3-10 on page 130 has a high third quartile. This occurs because share (or 'semi') elasticities (denoted by the superscript, w) are obtained by partial differentiation of the EASI demand equations with respect to income as in (Lewbel and Pendakur 2009, at p.835) or (Pendakur 2009, Eq. (23)): [21]

$$\hat{\epsilon}^{w,j}_{x,i} = \left.\frac{\partial w_j(\mathbf{p}, u, \mathbf{z})}{\partial u}\right|_i = \sum_{r=1}^{R} \hat{\beta}^{j}_{y,r} y_i^{r-1} \tag{3.11}$$

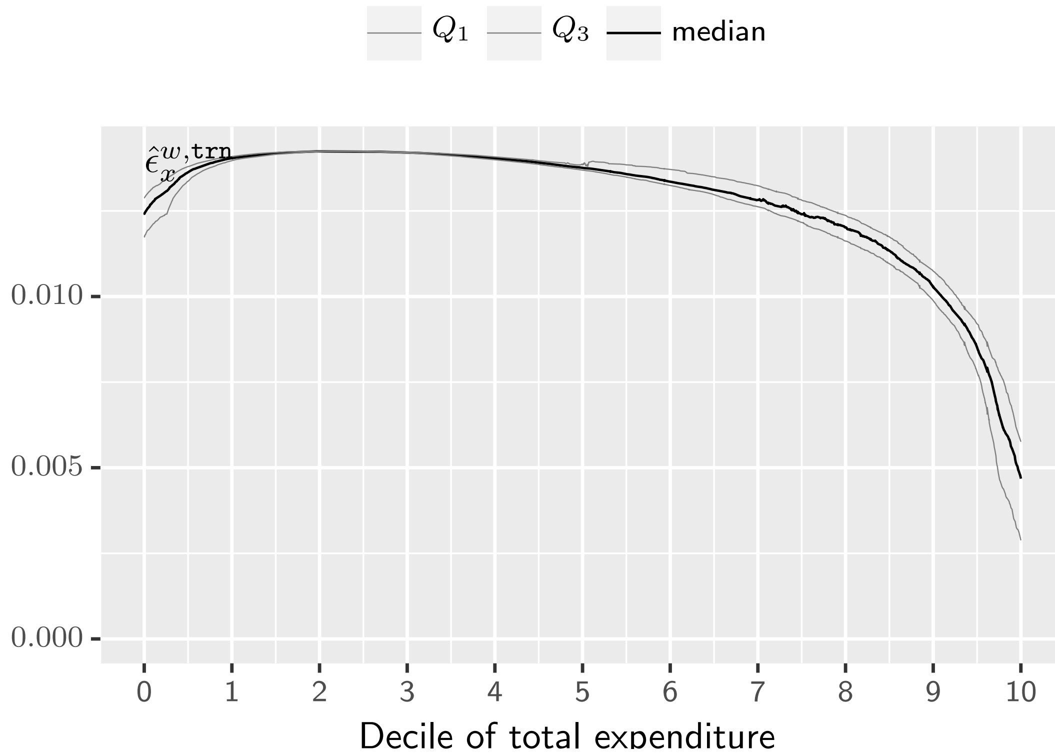
[21]Note that the current work does not explore interactions of prices with the demographics, or income with prices, and thus the corresponding terms from the cited equations are omitted.

To obtain the expenditure elasticity, the share elasticity is divided by the budget share w_i^j and added to 1. As the budget shares are very low in certain categories for some households, the low denominator leads to a high value, and so the interquartile range of the expenditure elasticity can be large: 1.26 to 2.78 at its peak, and 1.20 to 1.57 at median income. In contrast, the interquartile range of $\hat{\epsilon}_{x,i}^{w,j}$ only reflects differences in y due to variation in price levels across households with similar total expenditure, x, and so is smaller: 0.0142 to 0.0142 at peak share elasticity, and 0.0137 to 0.0139 at median income.

Lastly, highly elastic demand for transport is in contrast to other categories of consumption, shown in Figure 3-11 on page 131 as both share and quantity elasticities. As suggested by its declining Engel curve, the expenditure elasticity of `food` is below unity across the range of incomes (equivalently, its share elasticity is everywhere negative); on the other hand, `hou` expenditure is inelastic at lower incomes but becomes elastic at the seventh decile, reaching an elasticity of 1.45 at the highest incomes. While expenditure elasticities reflect the simple fact that households increase absolute expenditure in *all* categories as incomes rise, the share elasticities in Figure 3-11 reflect how households will alter their division of budget in response to an incremental rise in income. For the category of transport—as well as `clo`, `dur`, and `ed`—I find that this response is uniformly to increase the share of budget, while the budget shares for other categories of consumption fall over part or all of the income range.

3.6.3 Variation in transport budget shares

Changes in income, via the flexible income elasticity of expenditure, only partly of the difference in the transport budget share across households. The demand systems estimated here also allow the budget shares to be influenced by measures of local conditions, which are associated with budget share shifts of similar magnitudes to



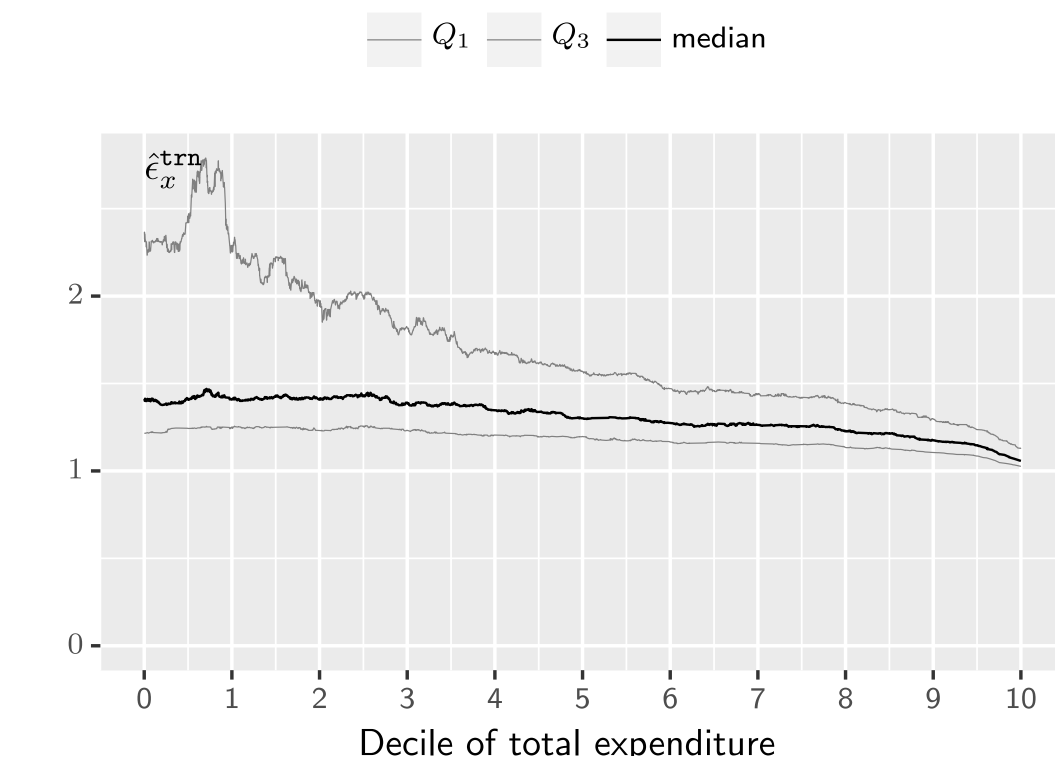


Figure 3-10: Share (top) and expenditure (bottom) elasticity of `trn` expenditure with respect to income, model `y5+hh+city`

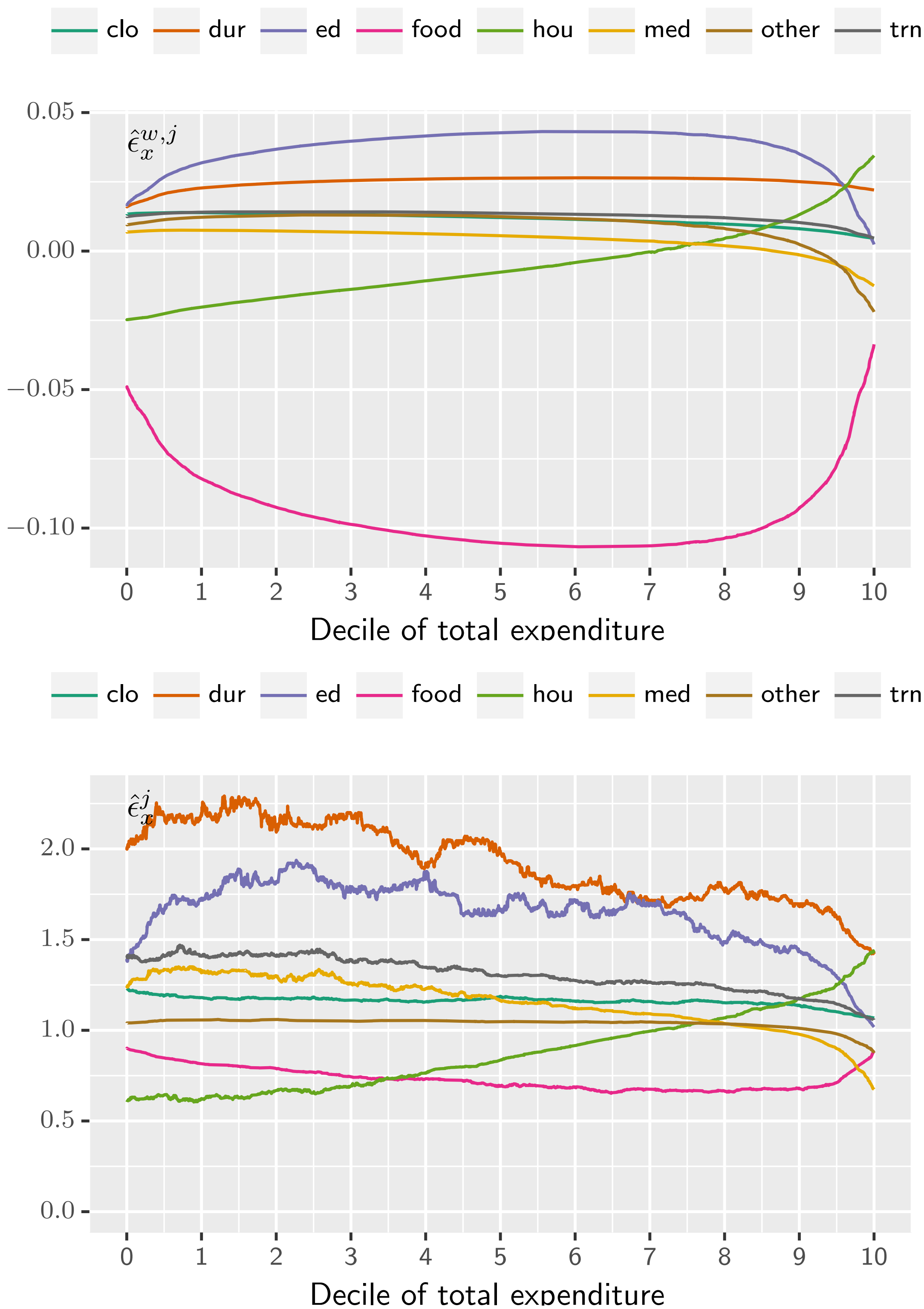


Figure 3-11: Share (top) and expenditure (bottom) elasticity of expenditure in all categories with respect to income, model `y5+hh+city`

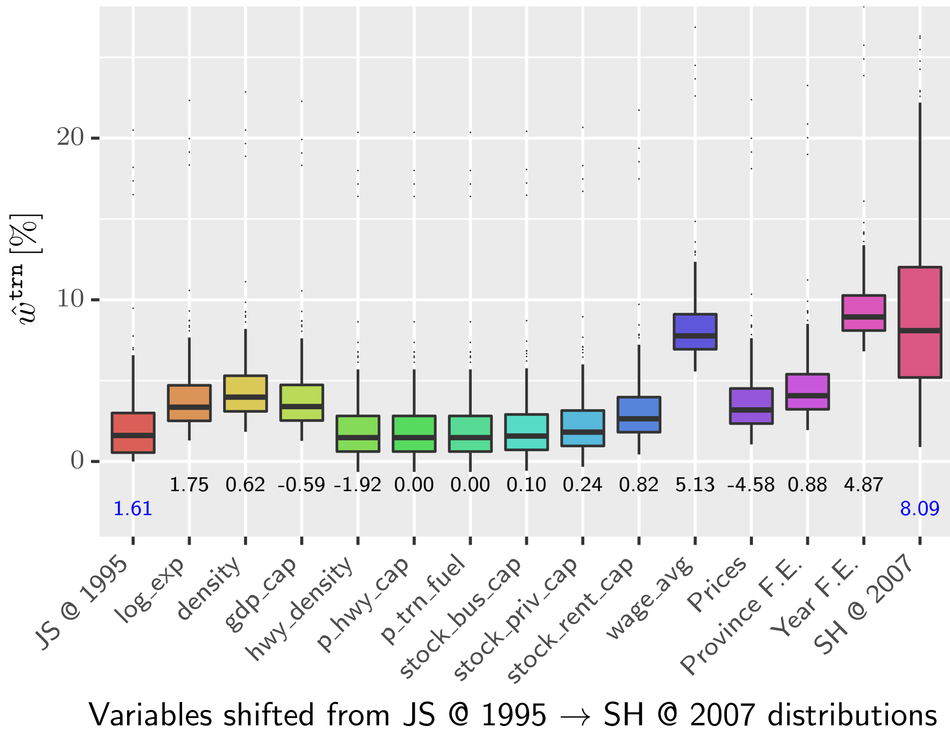


Figure 3-12: Shifts in budget share. Far left and right: distribution of the transport budget share for CHIP observations from cities in Jiangsu (JS) province, 1995; and Shanghai (SH), 2007, respectively, with medians in blue. Middle bars: hypothetical budget shares for the JS households as each successive variable is rescaled and shifted from its observed distribution to the SH 2007 distribution, with *changes* in the median in black relative to the adjacent bar to the left.

that caused by income. For selected households, for instance, only about 27 % of the difference in transport budget shares can be tied to increased income, with the remainder attributable to prices, local conditions, household-level demographics, and unobserved province- and year-specific attributes. Though comprehensive forecasts are beyond the scope of the current work, in order to illustrate the relative magnitude of the income and other effects, Figure 3-12 presents a comparison between two subsets of observations, using parameters from model `y6+hh+city`. I select, for the purpose of illustration, the relatively low-income households (median total expenditure: 9500 RMB) from Jiangsu (JS) province in the 1995 CHIP wave, and much higher-income households from Shanghai (SH) in the 2007 wave (median total expenditure: 39 000 RMB). Then, for successive variables z_t (first `log_exp`, then `density`, etc.), I shift these independent quantities, as in Equation (3.12), from the actual, observed values for the Jiangsu households, to new values z'_t that match the distribution of the same variable across the Shanghai households.

$$z'_{t,i} = \bar{z}_{t,SH2007} - \sigma_{z,t,SH2007} \times \frac{z_{t,i} - \bar{z}_{t,JS1995}}{\sigma_{z,t,JS1995}} \tag{3.12}$$

where

$$\begin{aligned} z_{t,i} &= \text{original value of } z_t \text{ associated with obs. } i \\ z'_{t,i} &= \text{shifted value} \\ \bar{z}_{t,JS1995}, \bar{z}_{t,SH2007} &= \text{means across subsets of obs.} \\ \sigma_{z,t,JS1995}, \sigma_{z,t,SH2007} &= \text{standard deviations} \end{aligned}$$

These values are used with estimated parameters to predict the budget shares shown, re-adding original prediction residuals. The predicted shares provide a convenient illustration of the incremental effect of differences in context between the two sets of observations.

The two sets of observations selected have a difference in median $w^{\texttt{trn}}$ of 6.48 percentage points of total expenditure. The change in expenditure between these conditions is associated with 1.75 points, or 27 %, of this gap. Shanghai's measures of density are higher: population density is shifted from a mean of 0.530 to 3.25×10^3 pers./km^2, associated with an increase in median $w^{\texttt{trn}}$ of 0.62 percentage points. Highway network density shifts from 0.250 to 1.76 km/km^2; this, in contrast, is associated with a decrease in transport budget share by 1.92 points—larger in magnitude than the change caused by income.

Measures of local vehicle stocks per capita are all greater for the 2007 Shanghai households than for the 1995 Jiangsu households, and each of these shifts is accompanied by an increase in the transport budget shares: 0.10 percentage points with larger bus fleets ($9.80 \times 10^{-2} \rightarrow 0.820\,\text{veh.}/10^3\text{pers.}$); 0.24 points with a larger stock of private vehicles ($0.950 \rightarrow 29.7\,\text{veh.}/10^3\text{pers.}$), and 0.82 points with a greater number of for-hire vehicles such as taxis ($0.0300 \rightarrow 2.36\,\text{veh.}/10^3\text{pers.}$). Both the local level of GDP per capita ($3.90 \rightarrow 62 \times 10^3\,\text{RMB/pers.}$) and average wages ($4.60 \times 10^3 \rightarrow 45 \times 10^3\,\text{RMB/(pers. year)}$) shift by more than an order of magnitude between the local conditions in Jiangsu 1995 and in Shanghai 2007, reflecting the rapid economic growth taking place in this period. The former is associated with an *decrease* of 0.59 percentage points in the transport budget share, while the latter is associated with a large increase of 5.13 points.

Changes in relative prices for transport and other goods faced by the two groups cause a 4.58 point decrease in the Shanghai households' transport budget share compared to the Jiangsu households. Other unobserved, time-invariant, province-specific factors captured by the provincial fixed effects are associated with an 0.88 point increase; while unexplained, nation-wide, year-specific factors captured by the year fixed effects are associated with 4.87 points of increase. While regional passenger road traffic per capita in 2007 Shanghai is lower than in 1995 Jiangsu, and fuel prices higher, neither of these variables are associated with transport budget share changes. Differences in household-level demographics explain the remaining difference between the medians in the rightmost two bars of Figure 3-12.

As noted, changes in income only explain roughly one quarter of the change in transport budget share between the two sets of households selected for this comparison, meaning that other factors are responsible for the remainder. This result highlights the importance of contextual factors in the transport system in mediating the consumption behavior of households. However—as discussed in Section 3.5.1—these remaining relationships are identified at best as associations rather than causal effects, and must be interpreted with care. For instance, the parameter for the stock of private vehicles (`stock_priv_cap`), is likely to be biased because this stock is created and replenished by the same transport expenditures that are the dependent variable in the demand equation, and so the regressor may be endogenous; a similar concern applies to `p_hwy_cap`. Separately, the large change associated of `wage_avg` may proxy for another, unobserved, city-specific attribute with which it is closely correlated. While the inclusion of province-level fixed effects captures unobserved attributes that differ across *provinces*, the CHIP sampled multiple cities within each province surveyed (cf. Table A.2), so estimation may assign budget share differences due to unobserved differences across prefectures to `wage_avg` or other included regressors. This possibility is further supported by the lack (discussed further below) of significance in the estimates for these parameters.

3.6.4 Parameter estimates for local conditions

I therefore offer only brief comment on the parameter estimates for certain city-level variables, in order to illustrate differences across alternate specifications of the EASI demand system, and to demonstrate that these measures of local conditions are clearly tied to household budgeting overall—though their effects on $w^{\mathtt{trn}}$ in particular are not estimated with significance. Figure 3-13 on the following page shows both t- and F-test results for the variables `density` and `gdp_cap` in the equation for $w^{\mathtt{trn}}$, across a variety of model specifications, along with confidence intervals at the 1%, 5% and 10% levels, and corrected for clustering where appropriate.[22] For a single model, Table 3.11 on page 137 gives the same information, plus the t-significance of the parameters for the variables in the other categories of consumption, j. Finally the reader is referred again to Table 3.10 on page 125, bottom panel, which shows differences in t- and F-significance in models that omit either province or year fixed effects.

For population density, as with other variables, coefficients are estimated with

[22]Similar figures are given in Appendix 3.B and Appendix A for other variables.

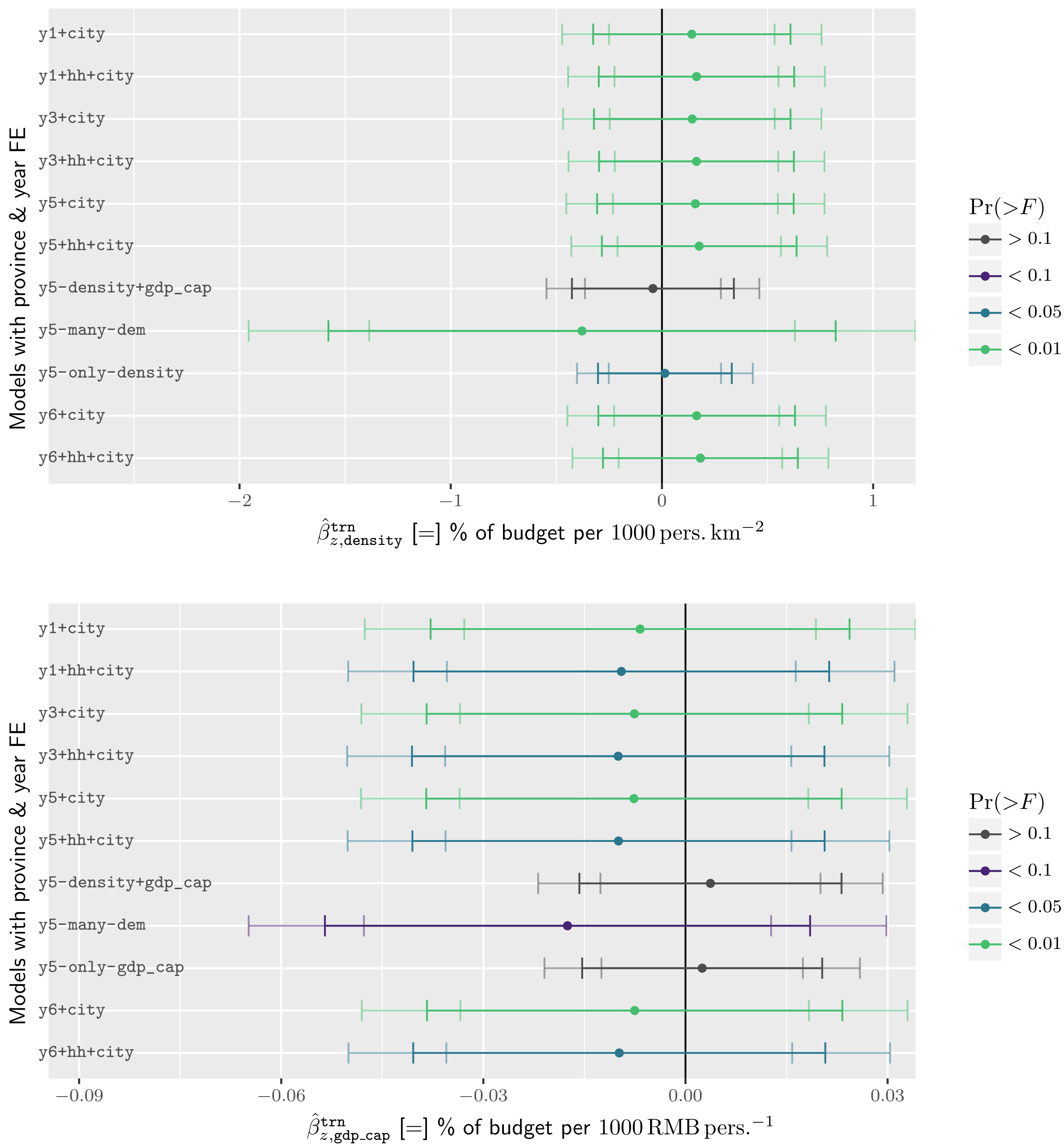


Figure 3-13: Estimates for coefficients on `density`, and `gdp_cap` in $w^{\mathtt{trn}}$, in multiple models. Whiskers give 1, 5, and 10% confidence intervals for individual parameter values using clustered standard errors. Colors indicate results of F-tests (Equation (3.10) on page 118) for association of variables with changes in budget allocation.

Table 3.11: Tests of significance for parameter influence on budget shares jointly (per Equation (3.9)) and individually (right columns), in model `y5+hh+city`. * = estimate significant at the 10% level, ** = 5%, *** = 1%.

	F	`food`	`clo`	`trn`	`hou`	`ed`	`dur`	`med`
`density`	4.78***	***	**					
`gdp_cap`	2.10**	*						
`hwy_density`	1.56			*	**			
`p_hwy_cap`	3.61***	**						
`p_trn_fuel`	0.293							
`stock_bus_cap`	1.12		**		*	*		
`stock_priv_cap`	1.54							*
`stock_rent_cap`	1.78*		*			*		
`wage_avg`	1.38		***		*	**		
`y5`	0.821	***	*		**	***	***	

significance in budget share equations for some categories, but not others. Per Table 3.11, in the model `y5+hh+city`, the estimate of $\beta^{\texttt{food}}_{z,\texttt{density}}$ is significant at the 1% level, and the estimate of $\beta^{\texttt{clo}}_{z,\texttt{density}}$ at the 5% level, while $\beta^{j}_{z,\texttt{density}}$ is not estimated with significance for other j. Figure 3-13 shows that, so long as both province and year fixed effects are included, no model has a $\beta^{\texttt{trn}}_{z,\texttt{density}}$ estimate that is significant at even the 10% level; however (column (4) of Table 3.10) if province fixed effects are omitted, the estimate is weakly significant at the 10% level.

The results of F tests described in Section 3.5.4 show that local population density is associated with changes in overall household budget allocation, as shown by the bar colors in Figure 3-13. Like the estimates of $\beta^{j}_{z,\texttt{density}}$ for other consumption categories, j, $\beta^{\texttt{trn}}_{z,\texttt{density}}$ has consistent sign and magnitude in models where the F test shows significance at the 1% level. In contrast, when this variable is the only city-level regressor (`y5-only-density`) or one of two (`y5-density+gdp_cap`), the coefficient estimate is lower and remains insignificant, and the F test of any influence on budget shares fails. In a model that deliberately introduces the somewhat collinear regressor `gdp_density` (`y5-many-dem`), the estimate for $\beta^{\texttt{trn}}_{z,\texttt{density}}$ is yet more negative.

Similar observations apply to the coefficients GDP per capita in CHIP households' locations, although the sign of $\beta^{\mathtt{trn}}_{z,\mathtt{gdp\_cap}}$ is negative. Restricted models (`y5-only-density` and `y5-density+gdp_cap`) and models with extra, collinear regressors (`y5-many-dem`) again produce biased estimates. When household-level demographics are omitted, the magnitude of $\beta^{\mathtt{trn}}_{z,\mathtt{gdp\_cap}}$ decreases while its F-statistic increases, since `educ` (and perhaps also `age`, `gender` and/or `single`) is correlated with `gdp_cap`.

In general, the interpretation of causal effects on transport behaviour awaits future work that is able to identify these effects clearly, for instance by investing in the collection of data suitable for instruments as discussed on page 114. Note, however, that the budget share systematization of demand, and the use of disaggregate household data, yield knowledge in a novel form that enriches the literature on the built environment. For instance (see again Table 3.1) Cervero and Murakami (2010) measured elasticities of a different dependent concept (VDT per vehicle per year), for different, aggregate units of analysis (US urbanized or statistical areas), and found that greater population density was associated with less vehicle travel, while greater road network density was associated with more VDT. In the EASI models, signs of $\beta^{\mathtt{trn}}_{z,\mathtt{density}}$ and $\beta^{\mathtt{trn}}_{z,\mathtt{hwy\_density}}$ are the opposite: positive and negative respectively. Yet this does not necessarily represent a contradiction; only the additional association that households spend a larger share of their budget on transport when they located in denser cities or districts, and a smaller share when road networks are denser. Per Gim (2012), the effects of population density in China may be different than what Cervero and Murakami find in the U.S. Or, higher VDT and associated costs may lead to a more than compensating decrease in consumption of other transport goods and services.

3.7 Discussion

Before concluding, I add some general observations about how the demand measure here relates to previously studied concepts (Section 3.7.1), and note some limitations on the analysis (Section 3.7.2).

3.7.1 Transport expenditure and its components

The models estimated in this chapter yield budget share curves and income elasticities for a category officially termed "transportation and communications." This category encompasses a basket of different goods and services with particular characteristics: some of the goods are durable, others not; some are complementary to others; and certain components may be substitutes and others complements. Using the U.S. Consumer Expenditure Survey, Choo et al. (2010) found that 'transport' and 'communications' are neither strict substitutes nor compliments, and that the relationship may be non-symmetric. The composition of the category varies with level of income and other factors: for instance, very poor households, or those not located near a city with an international airport, are unlikely to have international air trips as a component of their total transport expenditure. In the communications subcategory, while smartphones were not available during the years of the CHIP waves used here, cellular phone adoption and use grew dramatically.

Existing knowledge about households' transport expenditure (as reviewed in Section 3.2) centers on certain of these components—including fuel and vehicle purchases, and discrete choices of mode or numbers of trips. In order to relate mode- to budget shares, the costs of travel by different modes are required for conversion; to relate total transport expenditure to fuel and/or vehicle purchases, households' substitution behaviour across these goods is relevant. Work in data-rich contexts like the United States (e.g. Røed Larsen 2006) reveals that a household's income elasticity of demand for 'necessary' and 'luxury' transport goods can be quite different. Further, much of China's population in the time period studied, and likely a good share of the low-income households CHIP sample, used non-motorized transport for mobility; and walking and bicycling do not incur reportable variable costs.

To better understand what the present estimates imply about consumption of particular transport goods and services would require additional work to unpack the composition of Chinese households' transport expenditure, its relationship to non-monetary measures of mobility such as distance traveled by mode, and how these vary across incomes and the other measures of local conditions here.

3.7.2 Limitations

This work offers a foundation for further investigations of household-level demand for transport, linked to local conditions of the built environment, in flexible demand systems that, because they preserve rationality, can be readily linked to economy-wide models. Several of these possible are related to alternate or additional sources of data, or developing such links; those I describe in more detail in Chapter 5. Here, it is important to note two chief limitations in the present models.

Spatial variation in prices. The NBSC price indices provide changes relative to the previous year in the same region, but carry no information about spatial differences *across* regions. Biggeri et al. (2016), in providing an estimate of spatial price indexs (SPIs) for 2014, note that the most recent estimate in the literature was for 2002. They estimate general spatial price indices; that is, differences in purchasing power parities for a basket of representative goods or total consumption, rather than in the prices of distinct categories of goods. In contrast, C. Li and Gibson (2014) focus on one particular category of goods, namely housing, which is not easily traded across regions. No published source appears to cover spatial variation for all expenditure categories and years in the current data set, and at the same level of spatial resolution. Under the present treatment, these differences are partly absorbed by province and year fixed effects, but the addition of SPI information, if available, could improve the quality of estimates.

Treatment for endogeneity. The literatures on travel and the built environment, and on transport demand elasticities, have each identified and sought to address particular identification concerns: respectively, these are residential self-selection, and the simultaneity of demand and prices in markets, especially for transport fuels. I have demonstrated the EASI approach using readily-available, official statistics for urban variables; however, this introduces new concerns, as some of the available variables—in particular, `stock_priv_cap` and `p_hwy_cap`—are clearly endogenous with demand. Lacking viable instruments to control for these distinct forms of endogeneity, the city-level variables in the models function mainly as controls, and identification of causal effects of specific magnitude is precluded.

This limitation could be resolved by obtaining additional data from other sources, or assembling fresh data sets, with two goals: first, to include other measures of the built environment that are plausibly less endogenous with demand than the ones used here; and second, to provide suitable instruments for the data already collected. Cao et al. (2009) give some examples of the latter, although for transport-focused surveys and in the U.S. context; for instance, the instruments include shares of racial groups in population, a measure not likely to be available or useful as an instrument in the Chinese context. For the first goal, new data could be sought which better matches the conceptual quantities identified by the travel behaviour literature as having influence on trip choice. In particular, data on measures of supply in the public transport system, such as the track/route length, number of stations/stops, number of vehicles and/or lines, or fraction of population living within a certain distance of bus or rail public transit are desirable, because they reflect government policy decisions only partly predicated on current demand. These are likely to be less correlated with transport expenditures than measures of more granular, market-determined stocks such as the number of rental vehicles (`stock_rent_cap`), or measures of aggregate demand such as highway passenger-distance travelled (PDT) (`p_hwy_cap`). Such data would raise fewer concerns about endogeneity, and could be used to investigate the influence of public transport supply on transport expenditures.

3.7.3 Conclusions

This chapter has detailed a new application of a recently-developed, flexible demand system to data from publicly-available sources, in order to investigate the nature of transport demand in China and its relationship to local conditions.

The empirical results are a transport share of expenditure rising from 1.6 % to 7.5 % across the range of incomes, reflecting demand that is strongly elastic with respect to incomes among poorer households ($\epsilon_x^{\texttt{trn}} = 1.47$) and a declining income elasticity to $\epsilon_x^{\texttt{trn}} = 1.06$ at the highest incomes. While implying strong continued growth in the transport sector of China's economy as economic development leads to increased incomes and expenditure, these values are not as high as some prior estimates based on aggregate data analysed with less flexible demand systems.

The flexible, EASI demand systems yield two further conclusions: first, they reveal that third and fourth powers of implicit utility, y, a function of income, are significant in household budget allocations generally, and for transport in particular. Lower-order demand formulations therefore will not fully capture the detail of household responses to rising income. Second, the demand systems allow introduction of variables for city-level measures of transport system attributes, the built environment, and economic conditions, although no significant estimates were obtained with the present data for direct effects on transport spending. I find that these local conditions, along with prices and unobserved attributes of cities and specific years, are associated with the majority of variation in household transport budget share across selected conditions; only one quarter of the budget share difference between sets of selected households in 1995 and 2007 is explained by differences in income or total expenditure.

The findings build on previous efforts to estimate budget shares and elasticities of transport demand for Chinese households, shedding new light on how these descriptors of consumption are related to urban characteristics, and demonstrating how they can be estimated from disaggregate data from surveys not focused on travel or transportation.

One key application of such Engel curves is the projection of transport demand growth and energy consumption in Chinese cities—one that improves on the application of out-of-country, 20th-century trends, and reflects observed variation across and within cities. Models with city-level characteristics—combined with projections of demographics, income and its distribution, inflation, and planned transport infrastructure investment—enable exploration of correlations between future urbanization, economic growth, and aggregate household transport demand.

The demand systems I estimate could support policy analysis by allowing consistent estimation of the welfare impacts of transportation demand management (TDM) policies that either limit expenditure in particular categories, or affect the relative prices of transport goods and services. For instance, using microdata in Beijing and Shanghai, S. Li (2015) estimate welfare impacts of auction versus lottery distribution of vehicle license plates. My estimates of household transport expenditure based on country-wide data support similar analyses in other cities that may be considering such policies.

A key question for such projection and policy analysis, however, is whether EASI demands estimated on survey data like CHIP—with broad, national coverage, but only stratified sampling of a fraction of provinces in each wave—are valid for application in other cities countrywide. How sensitive are parameter values, and predicted budget shares, to the composition of the data used for calibration, and does the ability of the flexible system to capture complex income-demand relationships confer any advantage over more widely-used formulations such as AIDS? The following Chapter 4 takes up these questions.

References

Angrist, Joshua D. and Victor Lavy (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement*". In: *The Quarterly Journal of Economics* 114.2, pp. 533–575. DOI: 10.1162/003355399556061 (↪ p. 117).

Arzaghi, Mohammad and Jay Squalli (2015). "How price inelastic is demand for gasoline in fuel-subsidizing economies?" In: *Energy Economics* 50, pp. 117–124. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2015.04.009 (↪ pp. 85–87).

Biggeri, Luigi, Guido Ferrari, and Yanyun Zhao (2016). "Estimating Cross Province and Municipal City Price Level Differences in China: Some Experiments and Results". In: *Social Indicators Research*, pp. 1–19. ISSN: 1573-0921. DOI: 10.1007/s11205-016-1238-0 (↪ p. 140).

Burke, Paul J. and Shuhei Nishitateno (2013). "Gasoline prices, gasoline consumption, and new-vehicle fuel economy: Evidence for a large sample of countries". In: *Energy Economics* 36, pp. 363–370. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2012.09.008 (↪ pp. 85, 114).

Cameron, A. Colin and Douglas L. Miller (2015). "A Practitioner's Guide to Cluster-Robust Inference". In: *Journal of Human Resources* 50.2, pp. 317–372. DOI: 10.3368/jhr.50.2.317 (↪ p. 117).

Cao, Xinyu (Jason), Patricia L. Mokhtarian, and Susan L. Handy (2009). "Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings". In: *Transport Reviews* 29.3, pp. 359–395. DOI: 10.1080/01441640802539195 (↪ pp. 89, 141).

Caron, Justin, Valerie J. Karplus, and Giacomo A. Schwarz (2017). *Modeling the Income Dependence of Household Energy Consumption and its Implication for Climate Policy in China*. JP Report 314. Cambridge, MA, US: MIT Joint Program on the Science and Policy of Global Change. HDL: 1721.1/111814 (↪ pp. 94, 120, 196).

CEIC Data, ed. (2016). *China Premium Database*. URL: https://www.ceicdata.com/en/countries/china (visited on 12/14/2016) (↪ p. 106).

Cervero, Robert and Jin Murakami (2010). "Effects of Built Environments on Vehicle Miles Traveled: Evidence from 370 US Urbanized Areas". In: *Environment and Planning A: Economy and Space* 42.2, pp. 400–418. DOI: 10.1068/a4236 (↪

pp. 86, 89, 112, 138).
Chen, Y.-H. Henry, Sergey Paltsev, John M. Reilly, Jennifer F. Morris, and Mustafa H. Babiker (2015). *The MIT EPPA 6 Model: Economic Growth, Energy Use, and Food Consumption.* Tech. rep. 278. Cambridge, MA, US: MIT Joint Program on the Science and Policy of Global Change. HDL: 1721.1/95765 (↪ pp. 22, 91).
Cheung, Kui-Yin and Elspeth Thomson (2004). "The Demand for Gasoline in China: A Cointegration Analysis". In: *Journal of Applied Statistics* 31.5, pp. 533–544. DOI: 10.1080/0266476041000168183 7 (↪ pp. 86, 87).
Choo, Sangho, Taihyeong Lee, and Patricia Mokhtarian (2010). "Do Transportation and Communications Tend to Be Substitutes, Complements, or Neither?: U.S. Consumer Expenditures Perspective, 1984-2002". In: *Transportation Research Record. Journal of the Transportation Research Board* 2010, pp. 121–132. ISSN: 0361-1981. DOI: 10.3141/2010-14 (↪ p. 139).
Dahl, Carol A. (2012). "Measuring global gasoline and diesel price and income elasticities". In: *Energy Policy* 41. Modeling Transport (Energy) Demand and Policies, pp. 2–13. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2010.11.055 (↪ pp. 85, 86).
Dai, Hancheng, Toshihiko Masui, Yuzuru Matsuoka, and Shinichiro Fujimori (2012). "The impacts of China's household consumption expenditure patterns on energy demand and carbon emissions towards 2050". In: *Energy Policy* 50. Special Section: Past and Prospective Energy Transitions - Insights from History, pp. 736–750. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2012.08.023 (↪ pp. 94, 120).
Dargay, Joyce, Dermot Gately, and Martin Sommer (2007). "Vehicle ownership and income growth, worldwide: 1960-2030". In: *The Energy Journal* 28.4, pp. 143–170. DOI: 10.5547/ISSN0195-6574-EJ-Vol28-No4-7 (↪ pp. 24, 25, 112, 113).
Deaton, Angus and John Muellbauer (1980). "An Almost Ideal Demand System". In: *The American Economic Review* 70.3, pp. 312–326. ISSN: 00028282. URL: http://www.jstor.org/stable/1805222 (↪ pp. 92, 167).
Dulal, Hari Bansha, Gernot Brodnig, and Charity G. Onoriose (2011). "Climate change mitigation in the transport sector through urban planning: A review". In: *Habitat International* 35.3, pp. 494–500. ISSN: 0197-3975. DOI: 10.1016/j.habitatint.2011.02.001 (↪ p. 88).
Ewing, Reid and Robert Cervero (2010). "Travel and the Built Environment". In: *Journal of the American Planning Association* 76.3, pp. 265–294. DOI: 10.1080/01944361003766766 (↪ pp. 87, 88).
Fan, Shenggen, Eric J. Wailes, and Gail L. Cramer (1995). "Household Demand in Rural China: A Two-Stage LES-AIDS Model". In: *American Journal of Agricultural Economics* 77.1, pp. 54–62. DOI: 10.2307/1243888 (↪ p. 94).
Gim, Tae-Hyoung Tommy (2012). "A meta-analysis of the relationship between density and travel behavior". In: *Transportation* 39.3, pp. 491–519. ISSN: 1572-9435. DOI: 10.1007/s11116-011-9373-6 (↪ pp. 88, 138).
Goodwin, Phil, Joyce Dargay, and Mark Hanly (2004). "Elasticities of Road Traffic and Fuel Consumption with Respect to Price and Income: A Review". In: *Transport Reviews* 24.3, p. 275. DOI: 10.1080/0144164042000181725 (↪ pp. 85, 86).
Havranek, Tomas and Ondrej Kokes (2015). "Income elasticity of gasoline demand: A meta-analysis". In: *Energy Economics* 47, pp. 77–86. ISSN: 0140-9883. DOI:

10.1016/j.eneco.2014.11.004 (↪ pp. 85, 86).

Hoareau, Stéphane, Guy Lacroix, Mirella Hoareau, and Luca Tiberti (2012). *Exact Affine Stone Index Demand System in R: The EASI package*. URL: https://github.com/cran/easi/blob/master/inst/doc/easi.pdf (↪ pp. 115, 117, 126, 230).

Holz, Carsten A. (2004). "China's Statistical System in Transition: Challenges, Data Problems, and Institutional Innovations". In: *Review of Income and Wealth* 50.3, pp. 381–409. DOI: 10.1111/j.0034-6586.2004.00131.x (↪ pp. 47, 109, 165).

— (2013). "Chinese statistics: classification systems and data sources". In: *Eurasian Geography and Economics* 54.5–6, pp. 532–571. DOI: 10.1080/15387216.2014.908312 (↪ pp. 109, 165).

Hou, Quan and Si-Ming Li (2011). "Transport infrastructure development and changing spatial accessibility in the Greater Pearl River Delta, China, 1990–2020". In: *Journal of Transport Geography* 19.6. Special section on Alternative Travel futures, pp. 1350–1360. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2011.07.003 (↪ pp. 28, 102).

Huo, Hong, Kebin He, Michael Wang, and Zhiliang Yao (2011). "Vehicle technologies, fuel-economy policies, and fuel-consumption rates of Chinese vehicles". In: *Energy Policy (in print)*. DOI: 10.1016/j.enpol.2011.09.064 (↪ p. 94).

Jiang, Baichen and John Davis (2007). "Household food demand in rural China". In: *Applied Economics* 39.3, pp. 373–380. DOI: 10.1080/00036840500428013 (↪ p. 94).

Karathodorou, Niovi, Daniel J. Graham, and Robert B. Noland (2010). "Estimating the effect of urban density on fuel demand". In: *Energy Economics* 32.1, pp. 86–92. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2009.05.005 (↪ p. 88).

Kline, Rex B. (2012). "Assumptions of structural equation modeling". In: *Handbook of structural equation modeling*. Ed. by Rick H. Hoyle. New York, US: Guilford Press. Chap. 7, pp. 111–125. ISBN: 9781606230770 (↪ p. 89).

Labandeira, Xavier, José M. Labeaga, and Xiral López-Otero (2017). "A meta-analysis on the price elasticity of energy demand". In: *Energy Policy* 102, pp. 549–568. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2017.01.002 (↪ p. 85).

Lewbel, Arthur and Krishna Pendakur (2009). "Tricks with Hicks: The EASI Demand System." In: *American Economic Review* 99.3, pp. 827–863. ISSN: 0002-8282. DOI: 10.1257/aer.99.3.827 (↪ pp. 92, 95, 126, 128, 160, 163, 185, 196).

Li, Chao and John Gibson (2014). "Spatial Price Differences and Inequality in the People's Republic of China: Housing Market Evidence". In: *Asian Development Review* 31.1, pp. 92–120 (↪ p. 140).

Li, Lianyou, Ze Song, and Chao Ma (2015). "Engel curves and price elasticity in urban Chinese Households". In: *Economic Modelling* 44, pp. 236–242. ISSN: 0264-9993. DOI: 10.1016/j.econmod.2014.10.002 (↪ pp. 99, 117).

Li, Shanjun (2015). *Better Lucky Than Rich? Welfare Analysis of Automobile License Allocations in Beijing and Shanghai*. Working paper. DOI: 10.2139/ssrn.2349865 (↪ p. 143).

Li, Shi, Chuliang Luo, Zhong Wei, and Ximing Yue (2008). "Appendix: The 1995 and 2002 Household Surveys: Sampling Methods and Data Description". In:

Inequality and Public Policy in China. Ed. by Björn Gustafsson, Shi Li, and Terry Sicular. Cambridge University Press, pp. 337–353 (↪ p. 97).

Lin, Boqiang and Xiaoling Ouyang (2014). "A revisit of fossil-fuel subsidies in China: Challenges and opportunities for energy price reform". In: *Energy Conversion and Management* 82, pp. 124–134. ISSN: 0196-8904. DOI: 10.1016/j.enconman.2014.03.030 (↪ p. 102).

Lin, C.-Y. Cynthia and Jieyin (Jean) Zeng (2013). "The elasticity of demand for gasoline in China". In: *Energy Policy* 59.0, pp. 189–197. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2013.03.020 (↪ pp. 85, 86, 114, 127).

Luo, Chuliang, Shi Li, Terry Sicular, Quheng Deng, and Ximing Yue (2013). "Appendix I: The 2007 Household Surveys: Sampling Methods and Data Description". In: *Rising Inequality in China. Challenges to a Harmonious Society*. Ed. by Shi Li, Hiroshi Sato, and Terry Sicular. New York, NY, USA: Cambridge University Press, pp. 445–464. ISBN: 978-1-107-00291-3 (↪ p. 97).

Marchetti, Cesare (1994). "Anthropological invariants in travel behavior". In: *Technological Forecasting and Social Change* 47.1, pp. 75–88. ISSN: 0040-1625. DOI: 10.1016/0040-1625(94)90041-8 (↪ pp. 111, 113, 120).

McIntosh, James, Roman Trubka, Jeff Kenworthy, and Peter Newman (2014). "The role of urban form and transit in city car dependence: Analysis of 26 global cities from 1960 to 2000". In: *Transportation Research Part D: Transport and Environment* 33, pp. 95–110. ISSN: 1361-9209. DOI: 10.1016/j.trd.2014.08.013 (↪ pp. 88, 89).

McRae, Robert (1994). "Gasoline demand in developing Asian countries." In: *Energy Journal* 15.1, pp. 143–155. ISSN: 0195-6574. DOI: 10.5547/ISSN0195-6574-EJ-Vol15-No1-9 (↪ p. 86).

Moulton, Brent R. (1986). "Random group effects and the precision of regression estimates". In: *Journal of Econometrics* 32.3, pp. 385–397. ISSN: 0304-4076. DOI: 10.1016/0304-4076(86)90021-7 (↪ p. 117).

National Bureau of Statistics of China, ed. (2008). *China Statistical Yearbook*. 2008. Beijing, CN: China Statistics Press (↪ p. 109).

Newman, Peter G. and Jeffrey R. Kenworthy (1989). *Cities and Automobile Dependence. An International Sourcebook*. Brookfield, VT, US: Gower. ISBN: 0-566-07040-5 (↪ p. 88).

Oum, Tae Hoon, W. G. Waters, and Jong-Say Yong (1992). "Concepts of Price Elasticities of Transport Demand and Recent Empirical Estimates: An Interpretative Survey". In: *Journal of Transport Economics and Policy* 26.2, pp. 139–154. ISSN: 00225258. URL: http://www.jstor.org/stable/20052976 (↪ p. 85).

Parker, Steven (2018). *Transport Fuel Demand for Non-OECD Countries*. USAEE Working Paper 18-334. DOI: 10.2139/ssrn.3142665 (↪ pp. 87, 93).

Pendakur, Krishna (2009). "EASI made Easier". In: *Quantifying Consumer Preferences*. Ed. by Daniel Jonathan Slottje. 1st ed. Contributions to Economic Analysis 288. Bingley, UK: Emerald. Chap. 7, pp. 179–206. ISBN: 9781848553125. DOI: 10.1108/S0573-8555(2009)0000288010 (↪ pp. 95, 96, 117, 128, 194).

Rimmer, Maureen T. and Alan A. Powell (1996). "An implicitly additive demand system". In: *Applied Economics* 28.12, pp. 1613–1622. DOI: 10.1080/000368496327589

(↪ pp. 92, 167).
Røed Larsen, Erling (2006). "Distributional effects of environmental taxes on transportation: evidence from Engel curves in the United States". English. In: *Journal of Consumer Policy* 29.3, pp. 301–318. ISSN: 0168-7034. DOI: 10.1007/s10603-006-9009-y (↪ pp. 93, 121, 128, 139).
Sun, Chuanwang and Xiaoling Ouyang (2016). "Price and expenditure elasticities of residential energy demand during urbanization: An empirical analysis based on the household-level survey data in China". In: *Energy Policy* 88, pp. 56–63. ISSN: 0301-4215. DOI: 10.1016/j.enpol.2015.10.012 (↪ pp. 86, 94, 127).
Train, Kenneth E. (2009). *Discrete choice methods with simulation.* 2nd ed. New York City, NY, US: Cambridge University Press. 388 pp. ISBN: 9780521766555 (↪ p. 88).
van de Coevering, Paul and Tim Schwanen (2006). "Re-evaluating the impact of urban form on travel patternsin Europe and North-America". In: *Transport Policy* 13.3, pp. 229–239. ISSN: 0967-070X. DOI: 10.1016/j.tranpol.2005.10.001 (↪ p. 88).
Wang, H., P. Zhou, and D.Q. Zhou (2012). "An empirical study of direct rebound effect for passenger transport in urban China". In: *Energy Economics* 34.2, pp. 452–460. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2011.09.010 (↪ pp. 86, 94, 126, 127).
Yang, Zhenshan, Peng Jia, Weidong Liu, and Hongchun Yin (2017). "Car ownership and urban development in Chinese cities: A panel data analysis". In: *Journal of Transport Geography* 58, pp. 127–134. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2016.11.015 (↪ p. 94).
Zahavi, Yacov and Antti Talvitie (1980). "Regularities in Travel Time and Money Expenditures". In: *Transportation Research Record. Journal of the Transportation Research Board* 750, pp. 13–19. ISSN: 0361-1981 (↪ pp. 111, 121).
Zhao, Pengjun, Bin Lu, and Gert de Roo (2010). "Urban Expansion and Transportation: The Impact of Urban form on Commuting Patterns on the City Fringe of Beijing". In: *Environment and Planning A: Economy and Space* 42.10, pp. 2467–2486. DOI: 10.1068/a4350 (↪ p. 89).
— (2011). "The impact of urban growth on commuting patterns in a restructuring city: Evidence from Beijing". In: *Papers in Regional Science* 90.4, pp. 735–754. DOI: 10.1111/j.1435-5957.2010.00343.x (↪ p. 89).
Zhou, Ronggui, Liande Zhong, Nale Zhao, Jing Fang, Hua Chai, Jian Zhou, Wei Li, and Bing Li (2016). "The Development and Practice of China Highway Capacity Research". In: *Transportation Research Procedia* 15, pp. 14–25. ISSN: 2352-1465. DOI: 10.1016/j.trpro.2016.06.002 (↪ p. 106).

3.A Tables

Table 3.12: Estimated Engel curve parameters for model `y3+hh`. * = estimate significant at the 10% level, ** = 5%, *** = 1%.

	`food`	`clo`	`trn`	`hou`	`ed`	`dur`	`med`
				Powers of (implicit) utility			
$\hat{\beta}^j_{u,0}$	−30.858***	6.929***	10.780***	−6.748***	17.421***	10.483***	64.132***
	(5.951)	(2.414)	(1.593)	(2.248)	(3.377)	(2.273)	(2.571)
$\hat{\beta}^j_{u,1}$	44.031***	−2.697**	−3.831***	11.861***	−5.779***	−6.525***	−21.955***
	(2.818)	(1.143)	(0.754)	(1.064)	(1.599)	(1.077)	(1.217)
$\hat{\beta}^j_{u,2}$	−5.978***	0.504***	0.564***	−1.972***	0.705***	0.830***	2.602***
	(0.428)	(0.174)	(0.115)	(0.162)	(0.243)	(0.164)	(0.185)
$\hat{\beta}^j_{u,3}$	0.220***	−0.021***	−0.021***	0.091***	−0.017	−0.024***	−0.099***
	(0.020)	(0.008)	(0.005)	(0.007)	(0.011)	(0.008)	(0.009)
				Household-level variables			
$\hat{\beta}^j_{z,\texttt{age}}$	2.229***	−2.409***	−0.347***	−0.128	−0.223*	−0.212**	0.856***
	(0.230)	(0.093)	(0.061)	(0.087)	(0.130)	(0.088)	(0.099)
$\hat{\beta}^j_{z,\texttt{educ}}$	−4.872***	0.379***	0.373***	0.054	−0.075	−0.063	−0.551***
	(0.226)	(0.092)	(0.060)	(0.086)	(0.129)	(0.086)	(0.098)
$\hat{\beta}^j_{z,\texttt{gender}}$	−1.214***	0.684***	0.061	−0.118	0.401***	−0.198**	−0.364***
	(0.221)	(0.090)	(0.059)	(0.084)	(0.126)	(0.085)	(0.096)
$\hat{\beta}^j_{z,\texttt{single}}$	0.907**	−0.447**	0.244**	0.417**	−0.562**	0.653***	0.777***
	(0.432)	(0.175)	(0.116)	(0.163)	(0.245)	(0.165)	(0.187)
				Own- and cross-price elasticities			
$\hat{\beta}^j_{p,\texttt{food}}$	−5.421	−5.020**	−2.430	−0.584	3.185	3.417	5.469***
	(5.517)	(2.007)	(1.512)	(1.627)	(2.119)	(2.122)	(1.802)
$\hat{\beta}^j_{p,\texttt{clo}}$	−5.020**	5.495***	4.766***	1.067	0.380	8.249***	1.384
	(2.007)	(1.556)	(0.863)	(0.799)	(1.032)	(1.167)	(0.955)
$\hat{\beta}^j_{p,\texttt{trn}}$	−2.430	4.766***	3.572***	1.537**	−1.878**	0.467	0.741
	(1.512)	(0.863)	(1.028)	(0.634)	(0.756)	(1.001)	(0.689)
$\hat{\beta}^j_{p,\texttt{hou}}$	−0.584	1.067	1.537**	1.666*	4.185***	−0.319	−0.039
	(1.627)	(0.799)	(0.634)	(0.867)	(0.818)	(0.948)	(0.699)
$\hat{\beta}^j_{p,\texttt{ed}}$	3.185	0.380	−1.878**	4.185***	0.954	−1.171	2.891***
	(2.119)	(1.032)	(0.756)	(0.818)	(1.556)	(1.053)	(0.934)
$\hat{\beta}^j_{p,\texttt{dur}}$	3.417	8.249***	0.467	−0.319	−1.171	−5.204***	−9.865***
	(2.122)	(1.167)	(1.001)	(0.948)	(1.053)	(1.863)	(0.943)
$\hat{\beta}^j_{p,\texttt{med}}$	5.469***	1.384	0.741	−0.039	2.891***	−9.865***	0.517
	(1.802)	(0.955)	(0.689)	(0.699)	(0.934)	(0.943)	(1.109)

Table 3.13: Estimated Engel curve parameters for model `y5-many-dem`. * = estimate significant at the 10% level, ** = 5%, *** = 1%.

	`food`	`clo`	`trn`	`hou`	`ed`	`dur`	`med`
			Powers of (implicit) utility				
$\hat{\beta}^j_{u,0}$	17.897 (18.176)	−20.192*** (7.216)	−1.394 (5.785)	26.622*** (7.649)	−4.398 (11.012)	−23.742*** (7.155)	79.289*** (8.720)
$\hat{\beta}^j_{u,1}$	2.903 (23.254)	21.536** (9.176)	10.300 (7.089)	−24.325** (10.041)	−2.604 (14.537)	47.643*** (8.695)	−31.756*** (11.118)
$\hat{\beta}^j_{u,2}$	−3.664 (8.565)	−7.594** (3.380)	−3.840 (2.611)	10.386*** (3.699)	4.297 (5.355)	−16.103*** (3.203)	6.304 (4.095)
$\hat{\beta}^j_{u,3}$	1.549 (1.357)	1.108** (0.535)	0.546 (0.414)	−1.666*** (0.586)	−1.291 (0.848)	2.198*** (0.507)	−0.687 (0.649)
$\hat{\beta}^j_{u,4}$	−0.192* (0.099)	−0.071* (0.039)	−0.032 (0.030)	0.112*** (0.043)	0.137** (0.062)	−0.133*** (0.037)	0.042 (0.047)
$\hat{\beta}^j_{u,5}$	0.007*** (0.003)	0.002 (0.001)	0.001 (0.001)	−0.003** (0.001)	−0.005*** (0.002)	0.003*** (0.001)	−0.001 (0.001)
			Household-level variables				
$\hat{\beta}^j_{z,\texttt{age}}$	1.746*** (0.268)	−1.894*** (0.106)	−0.430*** (0.082)	−0.480*** (0.116)	−0.473*** (0.168)	−0.000 (0.100)	1.098*** (0.128)
$\hat{\beta}^j_{z,\texttt{educ}}$	−4.469*** (0.252)	0.381*** (0.099)	0.392*** (0.077)	0.070 (0.109)	−0.425*** (0.157)	−0.072 (0.094)	−0.445*** (0.120)
$\hat{\beta}^j_{z,\texttt{gender}}$	−1.124*** (0.251)	0.571*** (0.099)	0.039 (0.076)	−0.189* (0.108)	0.347** (0.157)	−0.064 (0.094)	−0.373*** (0.120)
$\hat{\beta}^j_{z,\texttt{single}}$	0.561 (0.469)	−0.290 (0.185)	0.272* (0.143)	0.477** (0.203)	−0.261 (0.293)	0.496*** (0.175)	0.853*** (0.224)
			City-level variables				
$\hat{\beta}^j_{z,\texttt{density}}$	4.849** (2.421)	−0.526 (0.766)	−0.378 (0.613)	−3.462*** (0.979)	−0.718 (1.137)	0.693 (0.637)	−0.285 (0.797)
$\hat{\beta}^j_{z,\texttt{gdp\_cap}}$	−0.105 (0.069)	−0.057*** (0.021)	−0.018 (0.018)	0.003 (0.028)	0.012 (0.032)	−0.016 (0.019)	−0.031 (0.023)
$\hat{\beta}^j_{z,\texttt{gdp\_density}}$	−0.041 (0.038)	0.019 (0.012)	0.009 (0.010)	0.046*** (0.015)	0.017 (0.018)	−0.012 (0.010)	0.007 (0.013)
$\hat{\beta}^j_{z,\texttt{hwy\_density}}$	−0.772 (3.400)	−0.528 (1.106)	−0.607 (0.856)	−0.470 (1.362)	−0.068 (1.578)	−0.139 (0.891)	0.210 (1.122)
$\hat{\beta}^j_{z,\texttt{p\_hwy\_cap}}$	171.288*** (65.698)	22.178 (21.060)	7.103 (16.568)	3.590 (26.216)	−29.286 (30.503)	0.491 (17.474)	0.249 (21.759)
$\hat{\beta}^j_{z,\texttt{p\_trn\_fac}}$	−32.511 (47.703)	−7.878 (10.738)	0.906 (20.968)	5.684 (13.200)	−15.400 (14.525)	−11.585 (22.592)	−23.179 (18.450)
$\hat{\beta}^j_{z,\texttt{p\_trn\_fuel}}$	30.977 (57.030)	−13.134 (16.919)	−27.462 (19.159)	−4.278 (17.997)	−0.486 (18.629)	39.316* (20.436)	9.886 (19.767)
$\hat{\beta}^j_{z,\texttt{p\_trn\_ic}}$	73.036 (45.964)	28.878** (12.756)	−30.931** (13.230)	−6.797 (14.848)	19.097 (16.587)	32.229** (14.045)	8.120 (14.542)
$\hat{\beta}^j_{z,\texttt{p\_trn\_maint}}$	−29.935 (32.905)	−14.723* (8.390)	−4.642 (12.746)	−9.809 (10.004)	10.262 (10.760)	10.128 (13.639)	−16.595 (11.856)
$\hat{\beta}^j_{z,\texttt{p\_trn\_pt}}$	−32.543 (21.033)	−21.554*** (6.780)	−2.976 (5.567)	−15.946* (8.423)	12.474 (9.529)	0.055 (5.919)	0.477 (6.943)
$\hat{\beta}^j_{z,\texttt{stock\_bus\_cap}}$	1.636 (3.399)	−2.422** (1.079)	−0.287 (0.847)	1.684 (1.387)	−3.165** (1.582)	0.402 (0.887)	0.411 (1.115)
$\hat{\beta}^j_{z,\texttt{stock\_priv\_cap}}$	−0.025 (0.066)	0.026 (0.021)	0.005 (0.016)	−0.005 (0.027)	−0.027 (0.031)	−0.008 (0.017)	0.024 (0.022)
$\hat{\beta}^j_{z,\texttt{stock\_rent\_cap}}$	1.112 (1.361)	0.838** (0.426)	0.406 (0.344)	0.238 (0.549)	1.411** (0.640)	−0.312 (0.361)	0.163 (0.449)
$\hat{\beta}^j_{z,\texttt{wage\_avg}}$	−0.450 (0.853)	0.399 (0.274)	0.025 (0.273)	−0.377 (0.284)	0.698** (0.314)	−0.188 (0.297)	−0.489* (0.291)

Table 3.14: Estimated Engel curve parameters for model `y5+hh+city`. * = estimate significant at the 10% level, ** = 5%, *** = 1%.

	`food`	`clo`	`trn`	`hou`	`ed`	`dur`	`med`
Powers of (implicit) utility							
$\hat{\beta}^j_{u,0}$	16.567	−26.705***	−2.856	26.650***	−4.504	−24.394***	69.651***
	(17.341)	(7.034)	(5.359)	(7.458)	(10.742)	(6.506)	(8.295)
$\hat{\beta}^j_{u,1}$	1.262	22.333**	9.132	−23.852**	−0.648	48.114***	−31.017***
	(23.261)	(9.188)	(7.083)	(10.063)	(14.523)	(8.685)	(11.099)
$\hat{\beta}^j_{u,2}$	−3.251	−7.903**	−3.381	10.296***	3.589	−16.368***	6.035
	(8.581)	(3.389)	(2.613)	(3.712)	(5.358)	(3.204)	(4.095)
$\hat{\beta}^j_{u,3}$	1.515	1.162**	0.471	−1.660***	−1.186	2.253***	−0.642
	(1.360)	(0.537)	(0.414)	(0.588)	(0.849)	(0.508)	(0.649)
$\hat{\beta}^j_{u,4}$	−0.191*	−0.076*	−0.027	0.112***	0.130**	−0.138***	0.039
	(0.099)	(0.039)	(0.030)	(0.043)	(0.062)	(0.037)	(0.047)
$\hat{\beta}^j_{u,5}$	0.007***	0.002*	0.001	−0.003**	−0.005***	0.003***	−0.001
	(0.003)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Household-level variables							
$\hat{\beta}^j_{z,\texttt{age}}$	1.835***	−1.908***	−0.431***	−0.524***	−0.477***	−0.005	1.084***
	(0.269)	(0.106)	(0.082)	(0.116)	(0.168)	(0.100)	(0.128)
$\hat{\beta}^j_{z,\texttt{educ}}$	−4.502***	0.385***	0.382***	0.076	−0.398**	−0.061	−0.436***
	(0.252)	(0.100)	(0.077)	(0.109)	(0.158)	(0.094)	(0.120)
$\hat{\beta}^j_{z,\texttt{gender}}$	−1.155***	0.539***	0.026	−0.216**	0.381**	−0.075	−0.370***
	(0.251)	(0.099)	(0.076)	(0.108)	(0.157)	(0.094)	(0.120)
$\hat{\beta}^j_{z,\texttt{single}}$	0.606	−0.300	0.283**	0.455**	−0.278	0.486***	0.837***
	(0.470)	(0.186)	(0.143)	(0.203)	(0.293)	(0.175)	(0.224)
City-level variables							
$\hat{\beta}^j_{z,\texttt{density}}$	2.673***	0.822**	0.176	−0.550	0.005	0.105	0.167
	(0.968)	(0.324)	(0.235)	(0.397)	(0.433)	(0.256)	(0.307)
$\hat{\beta}^j_{z,\texttt{gdp\_cap}}$	−0.121*	−0.034	−0.010	0.038	0.020	−0.024	−0.015
	(0.065)	(0.022)	(0.016)	(0.028)	(0.031)	(0.017)	(0.021)
$\hat{\beta}^j_{z,\texttt{hwy\_density}}$	1.840	−0.972	−1.268*	−2.293**	−0.758	0.300	0.023
	(2.812)	(0.981)	(0.699)	(1.122)	(1.205)	(0.734)	(0.888)
$\hat{\beta}^j_{z,\texttt{p\_hwy\_cap}}$	153.960**	17.420	6.189	−2.103	−30.098	10.795	−0.173
	(64.348)	(21.938)	(15.491)	(26.746)	(29.062)	(16.638)	(20.734)
$\hat{\beta}^j_{z,\texttt{p\_trn\_fuel}}$	16.152	−24.658	−1.176	14.785	−12.216	9.167	6.829
	(31.065)	(15.544)	(9.229)	(12.172)	(11.564)	(8.808)	(11.168)
$\hat{\beta}^j_{z,\texttt{stock\_bus\_cap}}$	−0.303	−2.633**	0.087	2.712*	−2.860*	−0.180	0.041
	(3.250)	(1.100)	(0.799)	(1.388)	(1.504)	(0.876)	(1.043)
$\hat{\beta}^j_{z,\texttt{stock\_priv\_cap}}$	−0.073	0.024	0.009	0.004	0.005	−0.004	0.029*
	(0.053)	(0.019)	(0.013)	(0.022)	(0.024)	(0.014)	(0.017)
$\hat{\beta}^j_{z,\texttt{stock\_rent\_cap}}$	1.823	0.787*	0.310	−0.284	1.037*	−0.132	0.158
	(1.267)	(0.423)	(0.306)	(0.537)	(0.590)	(0.337)	(0.406)
$\hat{\beta}^j_{z,\texttt{wage\_avg}}$	−0.305	0.667***	0.126	−0.398*	0.617**	−0.174	−0.090
	(0.583)	(0.247)	(0.161)	(0.233)	(0.247)	(0.160)	(0.192)

Table 3.15: Tests of significance for parameter influence on budget shares jointly (per Equation (3.9) on page 118) and individually (right columns), in `y5-many-dem`. * = estimate significant at the 10% level, ** = 5%, *** = 1%

	F	`food`	`clo`	`trn`	`hou`	`ed`	`dur`	`med`
`density`	3.52***	**			***			
`gdp_cap`	1.88*		***					
`gdp_density`	2.23**				***			
`hwy_density`	0.0918							
`p_hwy_cap`	4.02***	***						
`p_trn_fac`	0.593							
`p_trn_fuel`	0.842						*	
`p_trn_ic`	2.52**		**	**			**	
`p_trn_maint`	1.03		*					
`p_trn_pt`	1.88*		***		*			
`stock_bus_cap`	0.978		**			**		
`stock_priv_cap`	0.323							
`stock_rent_cap`	1.45		**			**		
`wage_avg`	1.04					**		*

Table 3.16: Specifications for `-only` models. • indicates inclusion of a variable (columns) in a model. Other columns: number of observations, N; powers of u included in the estimation, R.

	N	R	density	gdp_cap	gdp_density	hwy_density	p_hwy_cap	p_trn_fac	p_trn_fuel	p_trn_ic	p_trn_maint	p_trn_pt	stock_bus_cap	stock_priv_cap	stock_rent_cap	wage_avg	age	educ	gender	single
y5-many-dem	13357	5	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
y5-only-density	17689	5	•														•	•	•	•
y5-only-gdp_cap	14301	5		•													•	•	•	•
y5-only-gdp_density	14301	5			•												•	•	•	•
y5-only-hwy_density	17689	5				•											•	•	•	•
y5-only-p_hwy_cap	17689	5					•										•	•	•	•
y5-only-p_trn_fac	17689	5						•									•	•	•	•
y5-only-p_trn_fuel	17689	5							•								•	•	•	•
y5-only-p_trn_ic	17689	5								•							•	•	•	•
y5-only-p_trn_maint	17689	5									•						•	•	•	•
y5-only-p_trn_pt	17689	5										•					•	•	•	•
y5-only-stock_bus_cap	14024	5											•				•	•	•	•
y5-only-stock_priv_cap	17689	5												•			•	•	•	•
y5-only-stock_rent_cap	13357	5													•		•	•	•	•
y5-only-wage_avg	17689	5														•	•	•	•	•

3.B Figures

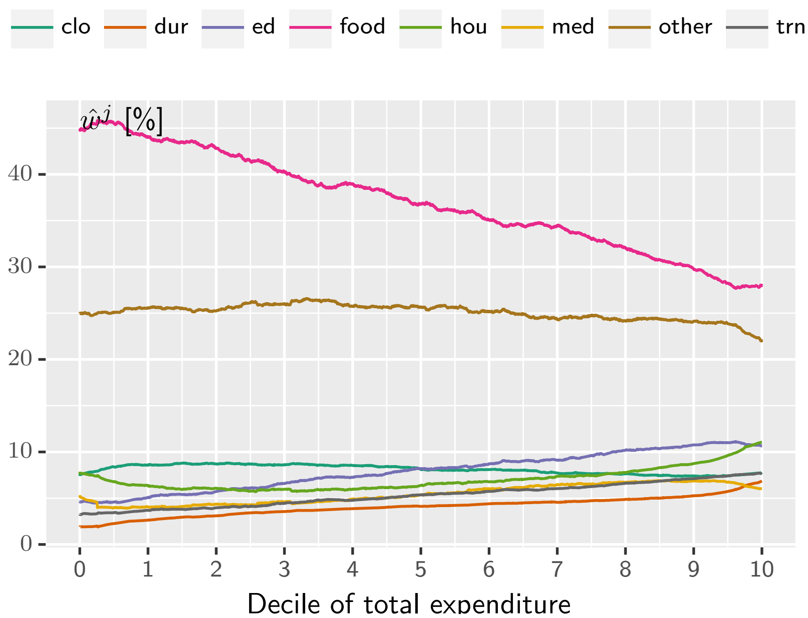


Figure 3-14: Fitted budget shares, model `y5+hh+city`

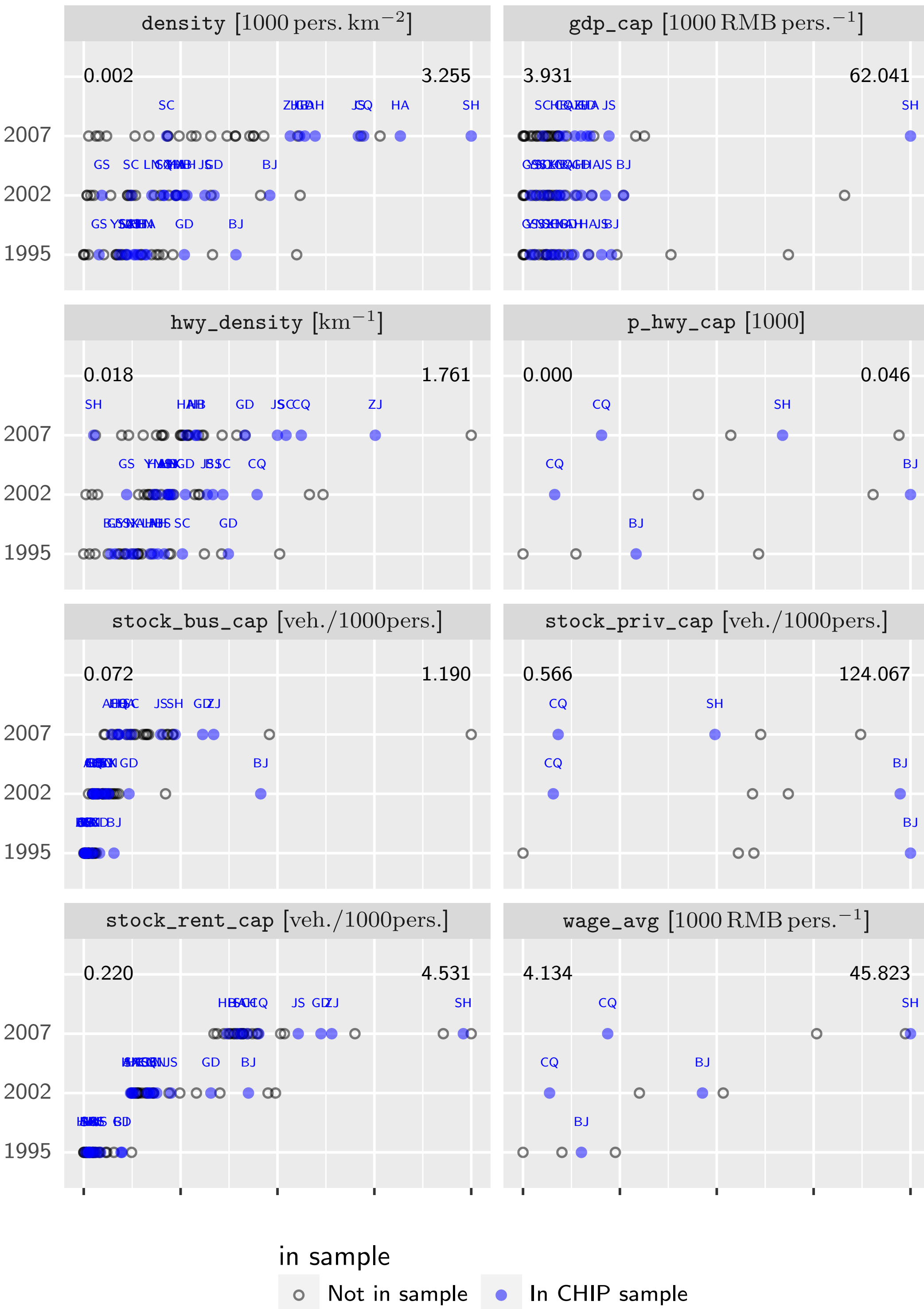


Figure 3-15: Descriptive statistics of provinces included in the CHIP sample.

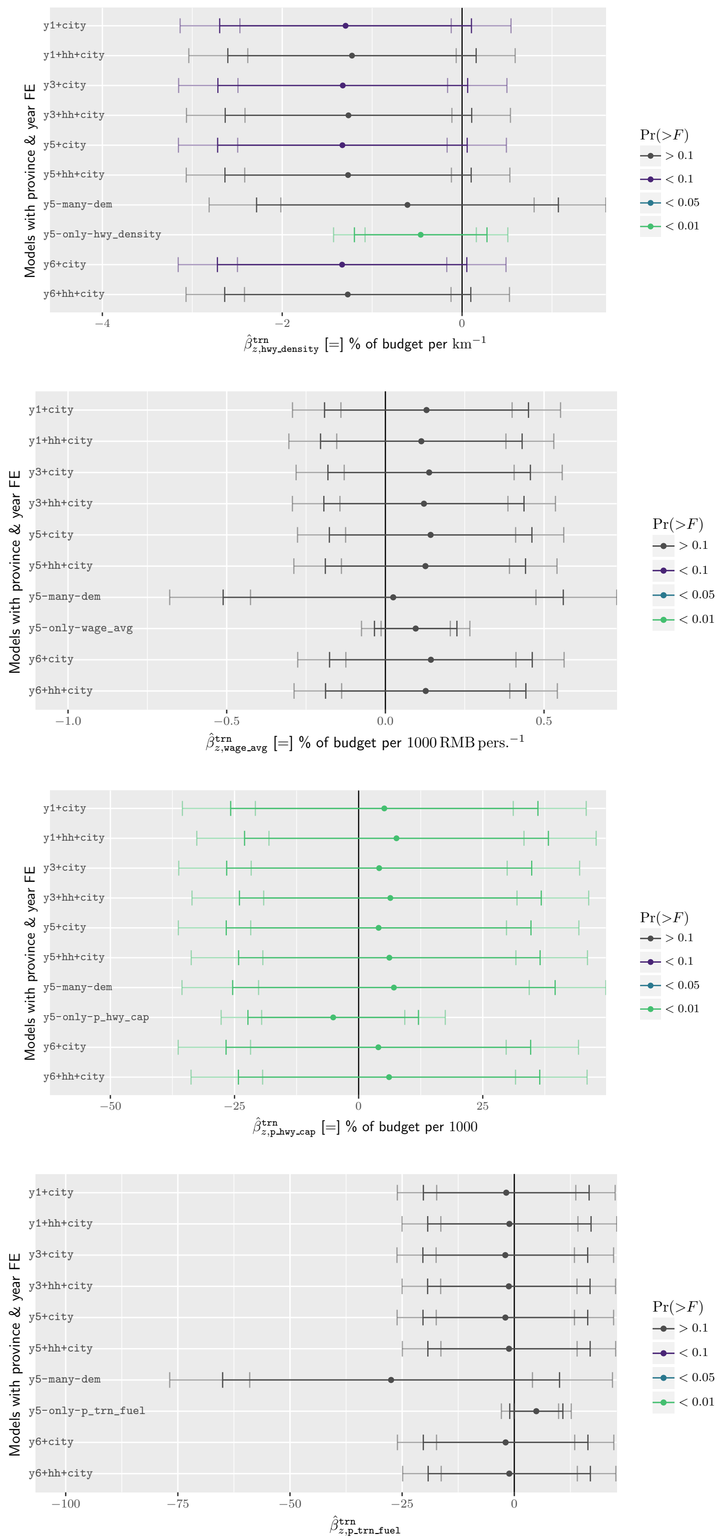


Figure 3-16: Estimates for coefficients on `hwy_density`, `wage_avg`, `p_hwy_cap`, and `p_trn_fuel` in $w^{\mathtt{trn}}$. Whiskers give 1, 5, and 10% confidence intervals for individual parameter values using clustered standard errors. Colors indicate results of F-tests (Equation (3.10) on page 118) for association of variables with changes in budget allocation.

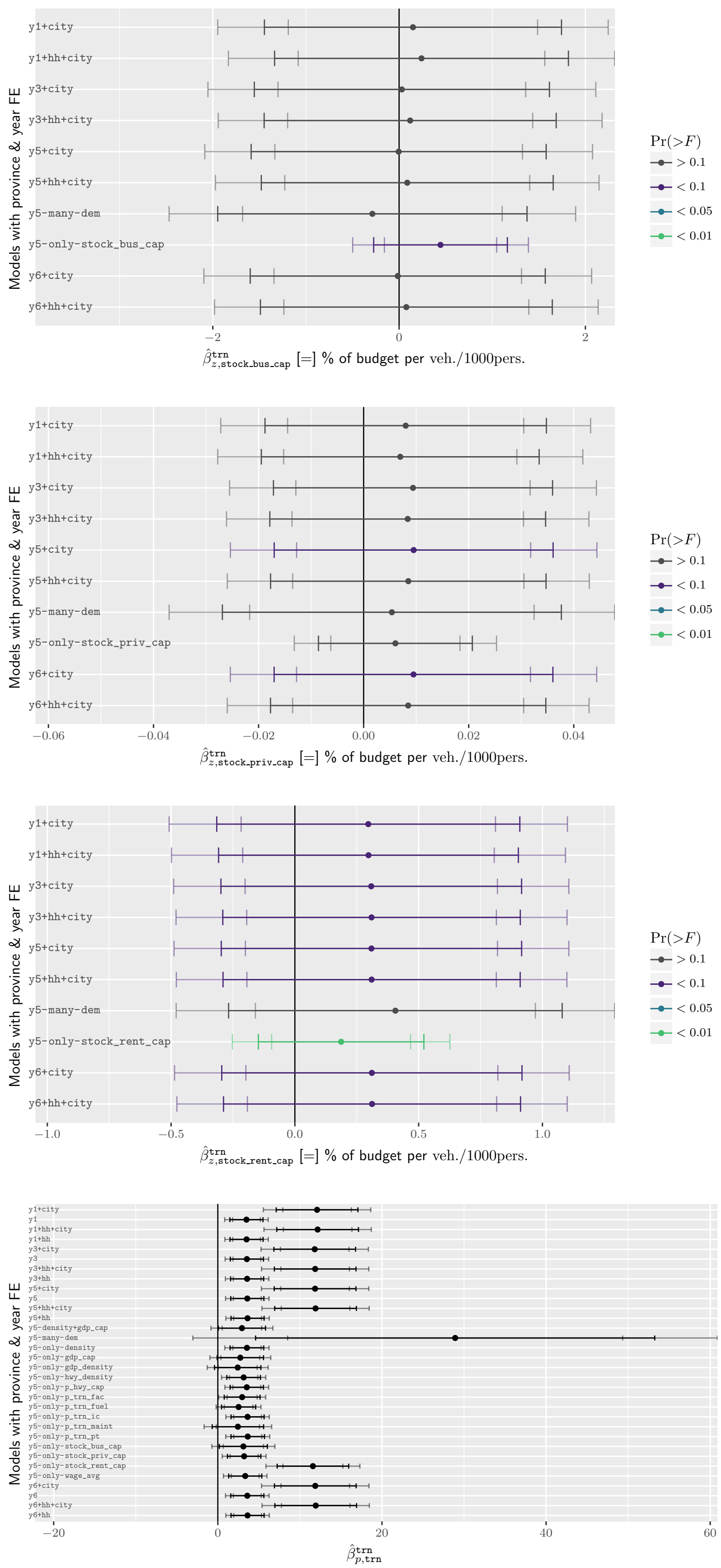


Figure 3-17: Estimates for coefficients on `stock_bus_cap`, `stock_priv_cap`, `stock_rent_cap`, and `ptrn` (the price index for the transportation category) in $w^{\texttt{trn}}$. See caption of Figure 3-16.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Validating flexible demand systems for Chinese household transport

Abstract

Empirical facts about transport demand and its relationship to economic growth are used to anticipate future activity and impacts, and to inform the design of transport policy. However, questions of external validity arise when translating these facts beyond the scope of supporting data. In this essay, I compare the recently-developed, Exact affine Stone index (EASI) demand system with the more widely used Almost Ideal demand system (AIDS), to examine if, how, and where its theoretical benefits translate to concrete advantages in projecting the transport behaviour of out-of-sample Chinese households.

Employing cross-validation techniques, I estimate both AIDS and a variety of EASI against fractions of a full data set, then test the performance of these models against the remainder. I find that simpler EASI specifications match or exceed the out-of-sample performance of AIDS models. Exploiting flexibility is not, however, without pitfalls: I show an accuracy trade-off when introducing policy-relevant covariates to models—especially where measures of local context are high; but also because relationships between these and travel behaviour may vary across provinces and cities. As well, controlling for unobserved, province-level confounders, while necessary for unbiased parameter estimates, yields models with increased error in out-of-sample prediction.

These results highlight the importance, in a diverse and rapidly changing country such as China, of data that sample the broadest possible range of city types and contexts; while the methods allow modelers to design and validate demand specifications for specific projection and assessment tasks.

4.1 Introduction

In transportation research, econometric methods and micro-scale data are used to infer and measure relationships between various measures of demand, and factors that influence it. Chapter 3 developed a new application of flexible demand systems to publicly-available survey data, in order to obtain empirical facts about Chinese households' transport behaviour. The Exact affine Stone index (EASI) equations used were originally developed in order to certain theoretical improvements over previous, less flexible forms (Lewbel and Pendakur 2009): namely allowing a more flexible relationship to income, prices, and demographic variables, while remaining utility-consistent in aggregation. Meanwhile, the China Household Income Project (CHIP) social survey data that formed the basis for estimation offered the advantage, compared to some prior research (Sections 3.2.1 and 3.2.2), of combining household resolution with nation-wide coverage.

Model-based assessment of projected future demand and the likely effect of policy alternatives is an important input to transport policy design that seeks to address externalities and impacts of demand (Section 1.2). In these processes, systems of demand equations can be used directly, or as a component of larger models and frameworks. Direct use occurs when the empirical facts from econometric inference—such as income elasticities, or the influence of certain demand drivers—are used to draw policy conclusions, or when simulated values of dependent variables are used to calculate outcomes under projected, policy, or counterfactual conditions. In contrast, within broader frameworks such as computable general equilibrium (CGE) and integrated assessment models (IAMs), demand equations do not fully determine outputs, but rather are linked, by economic or optimization logic, to representations of supply, technology, trade, and other concepts. Yet in both of these uses, the same question of portability arises: do the data reflect, and does the econometric method capture, relationships that are valid in the context(s) where policy insight is sought?

In this chapter, I compare the performance of alternative demand systems in terms of external validity. Specification of valid models is necessary in both the direct and framework uses in order to reliably project future demand and inform policy decisions. I introduce and demonstrate a cross-validation approach to testing alternate demand systems, specifications, and data coverage, in order to inform model design; in doing so, I also identify threats to external validity due to cross-province variation in travel behavior. I motivate by reviewing how demand systems are used as a component of transport model frameworks including IAMs (Section 4.2.1), and revisiting the mooted advantages of the EASI system over more widely-used formulations such as the Almost Ideal demand system (AIDS) generally (Section 4.2.2), and in potential CGE applications (Section 4.2.3). I then describe the geographical cross-validation approach (Section 4.3) in which AIDS and EASI models are estimated on subsets of household observations; used to predict demands for a withheld data segment; and then evaluated for performance. The results (Section 4.4) show that simple EASI formulations match, or exceed, the performance of AIDS (Section 4.4.1). EASI specifications with covariates describing local conditions and the built environment can worsen predictions (Section 4.4.2); a pitfall with multiple possible sources, one of which is a relationship between local context and travel behavior that varies across cities. Finally, I note a trade-off between controlling for unobserved confounders in order to obtain unbiased parameter estimates, and producing models that perform well out-of-sample (Section 4.4.3); this strongly suggests province-level heterogeneity in the way local conditions affect household spending. The chapter concludes with discussion (Section 4.5) of the implications for modeling demand in broader frameworks, and the characteristics of data that would improve portability.

4.2 Literature review

4.2.1 Demand formulations in integrated assessment models

IAMs and economy-wide models are used to project transportation activity, energy use and emissions globally, including for lower-income countries; and to compare the demand for, and impacts of, these activities with those from other sectors. CGE models calibrated on aggregate economic data are one type of model used for this purpose (e.g., Paltsev et al. 2016). A common CGE representation of household, or final demand is the constant elasticity of substitution (CES) production function, which yields homothetic preferences, or unitary income elasticity. This implies that the shares of different goods and services, in households' budgets, remain the same as economic growth causes incomes to rise; only variation in the relative prices in consumption categories will induce households to make budget adjustments.[1]

Karplus et al. (2013) relaxed this assumption by introducing the Linear expenditure system (LES) (i.e., Stone-Geary preferences) into the Economic Projection and Policy Analysis (EPPA) model, and used this formulation to study the adoption of plug-in hybrid electric vehicles (PHEVs) and other advanced technologies in the United States and other countries. The LES allows a parameter for a minimum, or 'subsistence,' level of consumption in each category of expenditure; varying this parameter allows representation of non-unitary income elasticity. Kishimoto (2012) employed this feature to study a period of rapid motorization in China, and future projections under reference and climate policy scenarios. Chen (2017) describes other efforts to introduce alternate formulations such as AIDS, "An implicit direct additive demand system" (AIDADS) and Constant difference of elasticities (CDE) demands into CGE models, evaluating calibration strategies for the latter.

[1] For a longer explanation of production function concepts, refer to Section 3.3.1.

4.2.2 Flexible demands: features and applications

In describing EASI demands, Lewbel and Pendakur (2009) listed their intended advantages, including (a) consistency under aggregation—such that demand systems can be used for welfare calculations—as well as (b) high degree in utility, u, and (c) the incorporation of demographics and controls through observation-level data, optionally interacted with prices and/or utility. In Chapter 3, I took advantage of (b) to estimate travel money budgets and estimate demand elasticities, showing that the demand of Chinese households for transportation & communications has significant third- and fourth-power terms; and of (c) to show that variables describing cities and transport systems have significant relationships to household budgets generally, and in some cases to the budget share of transportation, $w^{\texttt{trn}}$.

Beyond the domain of transport, EASI demand systems have been applied to empirical study of inequality in the effects on households of energy prices in Germany (Tovar Reaños and Wölfing 2018) and of food prices in Mexico (Wood et al. 2012). However, earlier demand systems are still in common use in empirical research. For instance, Fajgelbaum and Khandelwal (2016) use AIDS in a study of the distribution of gains from trade across consumers within countries. And EASI demand systems have not yet been connected to general equilibrium settings either at national or sub-national resolution, whether for the study of transport or other sectors; *cf.* again Chen (2017), and also Cogneau and Robilliard (2007), who list a small number of examples of (non-EASI) microsimulation models linked to CGE frameworks. Moreover, no study has explicitly compared the performance of EASI with the AIDS.

4.2.3 Potential use of flexible demands in CGE

Since welfare measures are commonly used in CGE policy assessment, the consistency feature of EASI demands raises the prospect of their use in CGE models. This would offer several benefits, which I elaborate here:

Higher-order functions of income. The flexible relationship between budget shares, w^j, and income, x (via implicit utility, y), would allow more precise computation of the welfare effects of transport-, climate- and other policies. Kishimoto (2012) found that climate policy sufficient to significantly impact the future course of motorization in China required high carbon prices and correspondingly large welfare and gross domestic product (GDP) impacts; but this result (and others) is grounded in the LES setting, rather than a flexible representation of household preferences. Higher-fidelity income- and price-elasticities from demand systems such as EASI would more accurately represent households' opportunities to substitute and change levels of consumption in order to meet policy constraints.

Incorporating variation in local conditions across city types. When projecting demand, the incorporation of household or city-level measures could capture effects raising or lowering demand that would otherwise be omitted. Exogenous projections (for instance, of population density and urbanization), stated policy targets (for instance, for the expansion of transport infrastructure), or other values for these variables could be incorporated explicitly in flexible demand formulations.

Work like that of S. Wang and J. Zhao (2018) suggests the value of incorporating city-level attributes in demand projection and policy analysis. Using time-series clustering on aggregate historical data including both broader contextual factors and transport-system characteristics, they identify four distinct groups of cities in China. Variables including the stock of rental vehicles, buses, and private light-duty vehicles (LDVs) per capita distinguished their "auto-oriented wealthy" cities from "low-density, medium-wealth, moderate mobility cities". Demand formulations that incorporate these variables would illuminate their effect on growing demand.

Providing policy counterfactuals. In line with #2, model-based policy assessment could be implemented through counterfactual values of exogenous variables. This offers the possibility of modeling policy with greater fidelity, as an alternative to converting non-price policy instruments to analogous price effects or quantity constraints that are compatible with currently-used demand formulations and model structures.

4.2.4 Partial-coverage survey data

In demonstrating the use of EASI demands in Chapter 3, I relied on a survey, CHIP, with a limited sample size. Since models used to project demand can be national or global in scope and use country-groups, countries or sub-national regions as their units of analysis, the question of portability or transferability arises: how well will demands represent aggregate expenditure when estimated on data from subset of the population? As complex engineering systems, the transport systems in various places may lead to heterogeneous behaviour, across dimensions that surveys may fail to span.

Like other low- and middle-income countries, China is still in the process of developing the institutions necessary to gather finer-resolution data on its economy and its sectors. This capability quantified by measures such as the World Bank Statistical Capacity Indicators (Stagars 2016). Holz (2004, 2013) documents the difficulties China has had in establishing a reliable system, including surveys of prices and industrial output, to measure GDP in a manner resistant to manipulation. Data capacities for sectors-specific measures, such as those in transport, lag capacities for more basic indicators of the necessities of life and economic performance. Where nation-wide surveys are conducted, data may be proprietary, or a difference in focus may mean that transport or energy questions are not included. The detailed transport- and energy surveys that *do* exist may have only local coverage in specific cities, and lack a longitudinal dimension, or information about their representativeness for other populations.[2]

CHIP, for instance, is a repeated-cross-section survey with a stratified sample. At the first stratum, the province level, between 9 and 12 units were selected in the waves used here. Consequently, any variation in transport expenditure particular to the remaining 19–21 provinces is not sampled. Other sources for China (e.g. the China General Social Survey (CGSS)) or for similar emerging economies (e.g. the National Sample Survey (India) (NSS) or General Household Survey (Nigeria) (GHS)) may also lack complete and uniform coverage, due to resource and capability constraints in data collection. One objective of the present work is to identify when inferences from such data can be safely used in on other contexts.

4.3 Methods, models and data

I test AIDS and EASI models for three purposes: first, to determine whether the benefits that motivated the development of EASI translate into improvements in projection accuracy in the context of transport; second, to study how this performance is tied to the inclusion of household- and city-level regressors in model specification; and third, to examine whether cross-province variation in transport system characteristics affects predictive performance. After briefly reviewing the methods of Chapter 3, this section covers the estimation of AIDS models (Section 4.3.1), the geographic cross-validation procedure (Section 4.3.2), and finally the error metric used to evaluate model performance under cross-validation (Section 4.3.3).

Chapter 3 described the data (Section 3.4) and methods (Sections 3.3.3 and 3.5.2) used to estimate various EASI models specifications. To reiterate, the CHIP observations consist of a non-panel stratified sample, with three waves in 1992, 2002, and

[2]In contrast, wealthy countries can finance frequent, public, large-scale and sophisticated data-collection efforts such as the U.S. National Household Travel Survey (NHTS).

2007 that covered a shifting subset of provinces, and of cities and districts at the lower strata within those provinces. Table 3.4 on page 100 gave the sample rates by province and year, while Table A.2 on page 217 gives the full list of units sampled. The data are augmented with official price series and measures of local conditions constructed from data in the "China Premium Database" published by CEIC Data (CEIC).

Chapter 3 found significant estimates for the transport budget share, $w^{\texttt{trn}}$, coefficients on y^3 and y^4, but not y^5 or higher powers. The analysis centers on a related set of models with this level of flexibility:

`y4` including only the powers of total expenditure $y^R, R \in 0 \ldots 4$. Versions with $R = 2$ and $R = 6$ are included for sensitivity checks.

`y4+hh` including household-level regressors for `age`, `educ`, `gender`, and `single`. (see Table 3.3 on page 99).

`y4+city` including nine covariates for city level conditions (see Table 3.6 on page 108).

`y4+hh+city` including both household- and city-level variables.

Each model is estimated with, or without, each of fixed effects by province (reference level: Beijing (BJ)) or year (reference level: 2002), for a total of four variations.

4.3.1 AIDS model estimation

I estimate the AIDS of Deaton and Muellbauer (1980) on the same data. AIDS is chosen because it is commonly used in CGE models (Dixon et al. 2013), as are related demand systems such as AIDADS (Rimmer and Powell 1996) and CDE (Chen 2017). I use the R software of Henningsen (2011) to perform the estimation.

A chief difference between AIDS and EASI is that the latter incorporates demographic controls as regressors in the budget share equations. (This was exploited in Chapter 3 to explore the relationship between regressors describing characteristics of urban and transport systems.) Consequently, estimation of AIDS does not require dropping observations with missing values for these variables where the CEIC source does not supply them; so the AIDS model is estimated on the entire CHIP sample of 18 624 households.

4.3.2 K-fold and geographic cross-validation

Cross-validation is a method of evaluating and comparing learning systems or algorithms by dividing the data into *training* and *validation* segments; the most common form is K-fold cross-validation (Refaeilzadeh et al. 2009). The procedure is as follows; Figure 4-1 on the next page shows it graphically:

1. Each observation, i, in the data set is randomly assigned to one of k segments $S_1 \ldots S_K$.
2. A model specification is estimated on the training segments $\{S_1, \ldots, S_{K-1}\}$.
3. The resulting parameter estimates are used to predict budget shares $\hat{w}_i^j$ for the withheld, validation segment, S_K.
4. Metrics of model fit and performance are computed; these are discussed below in Section 4.3.3.
5. Steps 2 through 4 are repeated, withholding each $S_k, k \in 1 \ldots K$ once.

For instance, if $K = 5$, so that each segment is 20% of the data, and the model is trained on 80% of the entire data set, and validated on the remaining 20%.

Cross-validation experiments are frequently used in transport research for methodological studies. For instance, Rengaraju and Arasan (1992) cross-validated a regression model of city-pair passenger air travel demand in India that was estimated on aggregate data. Nijkamp et al. (2004) use segmentation in experiments designed to test the relative performance of discrete choice and neural-network methods for modeling interregional freight transport flows in Europe. And M. Zhang et al. (2009) cross-validate partial least-squares regression models of demands for transportation energy in China. More broadly, researchers have studied relationships between in-

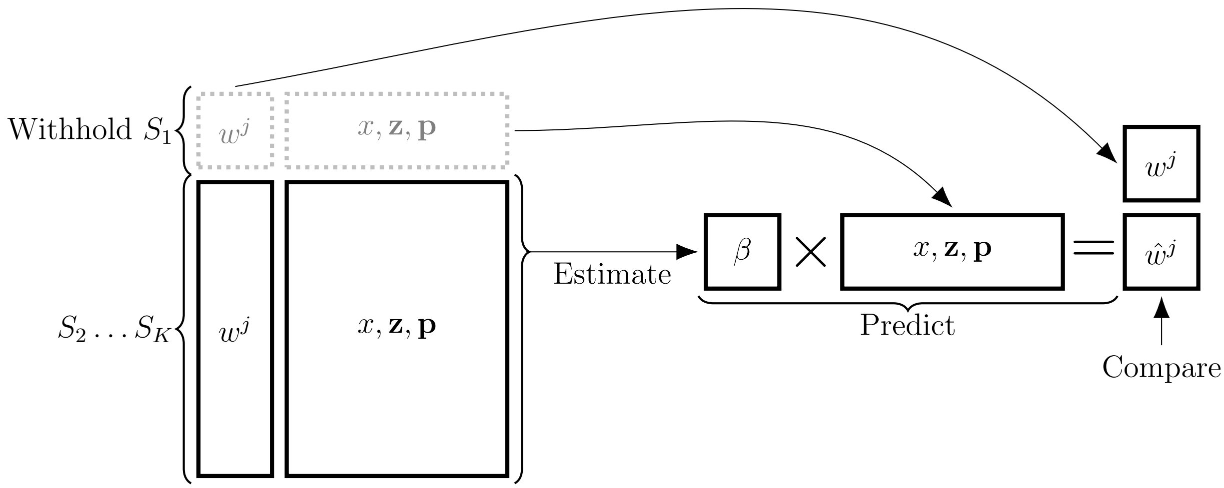


Figure 4-1: Schematic of cross-validation method. Training data from segments $S_2 \dots S_K$, including both independent $(\mathbf{p}, \mathbf{u}, \mathbf{z})$ and dependent (w^j) variables, are used to estimate model parameters β. Model parameters applied to independent variables from the withheld segment, S_1, to generate predicted $\hat{w}^j$, which are compared with w^j from the withheld segment. The process is repeated with S_2 withheld, etc.

and out-of-sample projections and the transferability of different methods between in-sample and out-of-sample contexts—for instance, Norwood et al. (2004) do this for crop yield models. Wenger and Olden (2012), while focusing on ecological models of population distributions, stress the importance of transferability testing when moving from samples to projecting broader populations; this is similar to my purpose in the present work.

The current chapter mainly employs a second type: geographic cross-validation. While the procedure is intuitive, applications are less common than of K-fold cross-validation, which can be applied to machine learning and other models trained on data that may have no clear spatial or geographical aspect. Instances include Wenger and Olden (2012), again; Pradhan (2010) uses cross-validation for remote-sensing models of landslides risk, comparing how models estimated on data from one location perform on another; and Farber and Páez (2007) use geographical subsetting to examine the performance of K-fold and Monte Carlo cross-validation in metaparameter selection for the specific method of geographically-weighted regression. Here, one motivation for geographical cross-validation is the stratified sampling method used by CHIP, in which only some of China's provinces are selected at the top stratum. In China—as in other low- and middle-income countries where transport activity growth is most rapid—resource constraints[3] make it difficult to field nationwide surveys, and thus research (including transport research) must sometimes grapple with partial-coverage data.

I make use of the same feature of the CHIP survey that enabled matching of the city-level regressors in Chapter 3 on page 81: the data contain the locations of individual households. The procedure is as above, except that in Step 1 the assignment is done so that observations in province p across CHIP waves are assigned non-randomly to a segment S_p. Because the CHIP 1995, 2002, and 2007 waves collectively surveyed respondents in 14 provinces, $P = 14$ groups are created. This has the effect of withholding entire provinces' worth of households from the data used for fitting model parameters. I label these segments with the GB/T 2260 alpha-2 code of the province; and label the validation models produced and their performance metrics with the name of the training segment that is withheld. For instance, "AH" (Anhui) refers to a model estimated on observations from all provinces *except* Anhui; and performance measures labelled "AH" refer to the predictions of (only) Anhui households' budget shares using this model.

As in Chapter 3, observations are pooled across years, and since each province was covered in one, two or all three of the CHIP waves (*cf.* again Table 3.4 on page 100), this implies that withholding each province's observations withholds data from only certain years. For instance, the Shanghai ("SH") validation segment includes households surveyed in 2007 but not 1995 or 2002; whereas the Sichuan ("SC") segment includes households surveyed in all three waves.

[3] including cost, but also the lack of experienced or professionalized surveyors.

4.3.3 Metrics and dimensions of model evaluation

In order to compare performance of models, I compute the root mean squared error (rmse) (Equation (4.3)) between the budget shares that the model predicts on the training set, and the true budget shares from the withheld observations. Since the rmse has the same units as the dependent variable w^j, the interpretation of a rmse of, for instance, 0.1 is that the sample standard deviation of model predictions in a particular category from the observed values is one tenth of households' total expenditure. For clarity, values are multiplied by a factor of 100 in the following tables and figures, so that rmse can be read as percentage points of total expenditure.

$$e_i^j = \hat{w}_i^j - w_i^j \quad \text{Prediction error for obs. } i \text{ in expenditure category } j \tag{4.1}$$

$$\frac{\sum_{i \in S_k} \left(e_i^j\right)^2}{|S_k|} \quad \text{Mean squared error for observations in } S_k \tag{4.2}$$

$$\sqrt{\frac{\sum_{i \in S_k} \left(e_i^j\right)^2}{|S_k|}} \quad \text{Root mean squared error for } S_k \tag{4.3}$$

Rather than give the aggregate rmse across all expenditure categories, j, I report and discuss values separately for each category.

Using these metrics, I make comparisons along three dimensions. First, models of different specifications are compared with one another. This *method-to-method* comparison reveals how functional form (when comparing AIDS to EASI models) or inclusion of certain regressors (when comparing EASI specifications) affects model performance—in particular, whether some models' base performance is more sensitive than others to the withholding of certain observations. Next, a model estimated without a validation segment S_k or S_p is compared to a model of the same specification estimated on the entire data set (the "base model"). This *segment-to-base* comparison informs whether the performance of the base model depends on the inclusion of observations from the withheld segment. Finally, models of the same specification estimated on different validation segments are compared with one another. This *segment-to-segment* comparison—of interest only for geographic cross-validation—identifies whether model results or performance are affected by province-to-province differences in the way households' preferences relate to city-level characteristics.

4.4 Cross-validation results and discussion

4.4.1 Comparing AIDS and EASI demands

The method-to-method comparison shows that certain EASI model specifications outperform the AIDS across provincial validation segments and budget share categories. Before highlighting the result, a brief guide to interpretation of Figure 4-2 on the next page. The top panel gives the rmse, for each provincial validation segment and budget category, of a simple EASI specification with 6 powers of implicit utility and neither household- nor city-level covariates, nor province or year fixed effects (models `y6`). Prediction errors for the transport budget share are lower than those for other categories, at 3.6 to 5.4 percentage points of expenditure across segments; provinces are sorted by this column. The prediction error of $w^{\texttt{trn}}$ in the base model is 3.7 points.[4] The provinces are sorted by this field; this means that when Chongqing households are withheld from the data used for estimation, the resulting model will predict their transport budget share with an error of 5.4 points of budget. Note that these are large errors, given the Engel curve result in Chapter 3 that the mean budget share varies from 1.6 % to 7.5 % across the range of incomes The highest prediction errors are seen for `food` and `other`; with reference to Figure 3-14 on page 154, note that the former two are also the categories with the largest budget share for households above the first income quintile, and also exhibit the largest changes in budget share

[4] This is, equivalently, the root sum of squared residuals for the model estimated on all data, with no segment withheld.

	clo	dur	ed	food	hou	med	other	trn
CQ	5.9	5.3	11.3	14.0	6.1	7.6	15.0	5.4
SH	4.6	5.2	9.8	11.6	5.5	5.2	13.9	5.3
ZJ	5.3	5.0	8.9	14.9	7.7	7.7	16.8	4.9
GD	5.9	6.0	8.9	15.9	7.3	5.9	11.0	4.9
HA	7.4	6.0	9.0	15.3	5.9	7.9	11.7	4.8
BJ	6.3	6.2	9.3	13.6	6.0	7.0	15.0	4.6
YN	6.3	5.8	8.6	12.6	5.7	5.8	10.2	4.6
LN	6.5	5.2	9.0	12.8	5.1	6.0	14.5	4.5
SX	8.1	5.9	9.4	14.6	5.2	6.9	12.4	4.4
HB	6.9	5.9	9.4	13.5	5.3	6.2	10.6	4.4
SC	5.8	6.2	8.1	12.9	5.1	6.5	10.9	4.2
AH	5.7	5.1	7.7	14.0	6.1	6.3	11.8	4.1
JS	5.9	5.8	8.7	14.2	5.1	6.7	10.9	4.1
GS	7.2	6.1	8.2	13.0	5.4	6.2	9.0	3.6

	clo	dur	ed	food	hou	med	other	trn
CQ	-3.9	-4.3	-4.0	-11.1	0.7	-0.1	-33.9	-8.2
SH	-1.7	-0.2	-0.8	-3.7	1.4	1.1	-1.5	-3.0
ZJ	0.3	0.5	1.4	-1.5	-2.1	0.6	11.5	3.3
GD	10.7	-2.1	0.5	-9.7	-1.0	2.1	-40.0	-8.4
HA	-6.8	-2.2	5.2	2.6	0.9	2.1	-34.5	-8.4
BJ	-6.7	-2.0	-10.0	-8.3	39.0	-3.7	-27.5	-11.3
YN	-2.5	0.3	-3.0	-25.7	17.7	-2.9	-48.4	-8.3
LN	-6.3	-0.2	-2.5	-9.2	0.0	-6.9	-26.0	-3.9
SX	-8.0	-2.8	-5.2	14.2	0.7	-0.1	-33.4	-3.4
HB	-5.4	-2.6	-4.1	-8.8	-2.7	-1.3	-45.7	-6.5
SC	-0.3	-3.1	-4.5	-14.4	-2.4	-4.2	-37.3	-3.9
AH	-4.9	-3.1	-1.1	-12.9	-3.9	-3.8	-35.9	-8.3
JS	-0.3	-2.3	-2.8	-12.7	-1.3	-1.1	-39.8	-1.7
GS	-8.9	-4.6	-5.1	-3.4	-4.9	-3.3	-53.9	-10.2

Figure 4-2: Root mean squared error for geographical cross-validation of model `y6` without fixed effects. Top: rmse comparing predicted expenditure shares versus observations, in each provincial segment, S_p (rows) and category, j (columns). Yellow indicates higher rmse. Bottom: percent improvement in rmse of `y6` relative to `aids`. Blue color indicates that the EASI model performs better; red, worse. See Figure 4-1 on page 169.

from the lowest to highest income.

The bottom panel of Figure 4-2 gives the percent change in prediction error for EASI model `y4` over the AIDS. In almost all provincial segment/category combinations, the flexible system produces small reductions in prediction error; however, there are others where the errors increase. For instance, $w^{\texttt{hou}}$ for Beijing households is predicted with a 39 % larger error (or 1.7 points of budget) by the EASI model when compared to the AIDS. The changes in transport budget share error are in the narrow range of $-11\,\%$ to $+3.3\,\%$, or -0.6 to $+0.2$ points of budget; a small but non-negligible amount compared to the overall prediction error. Finally, prediction errors for the category of `other` consumption are decreased significantly (up to 53 %) by the flexible model in almost all validation segment.

This modeling implication of this result is that simply-specified EASI models, as intended, can form a drop-in replacement for the AIDS. When estimated with 'missing' data from some provinces, AIDS models' predictive performance on the withheld set suffers, whereas the flexible system is not as strongly affected, as it capturing the significant, higher powers of variation in the relationship between income and transport budget share. Since the model `y6` does not use household-level regressors from CHIP responses (age, gender, education and marital status of the household head), there is no need to collect or synthesize these values when the estimated model is applied out-of-sample.

4.4.2 EASI model specifications and regressors

Flexible functional forms offer modelers a choice of many possible specifications, in combination with available data. Testing the performance of these under cross-validation, I show that these options present trade-offs between accuracy and the inclusion of policy-relevant factors in demand system specifications, because of underlying heterogeneity across Chinese cities in the way that such factors influence travel behaviour.

This is illustrated by a two-stage comparison, using a base model with 4 powers of implicit utility and year fixed effects (`y4`); the root sum of squared residuals (rssr, i.e. for all data) is 3.6 points of total expenditure. When four, household-level demographic variables are added, forming model `y4+hh`, the overall rssr remains the same. The model performs slightly better under cross-validation (Figure 4-3 on the next page, top panel): the demand system exploits the demographic information to produce negligible but consistent reductions in `clo` and `food` prediction errors, and reductions with some exceptions in other categories.

However, when city-level indicators of local conditions are added, forming model `y4+hh+city`, predictions for certain categories of expenditure become highly sensitive to the data used for estimation (Figure 4-3 on the following page, bottom panel). For instance, when Shanghai observations are withheld, then the resulting coefficients predict those households' `clo` expenditures with an rmse 330 % higher (or 14.9 points of total expenditure larger) than the model lacking city variables. These sensitivities are smaller for the consumption categories of `dur`, `hou`, `med`; for `trn`, aside from Jiangsu at 18 %, predictions errors change by $-4.3\,\%$ to $+3.3\,\%$ (-0.2 to $+0.1$ points).

The difference between the performance impacts of adding the household- and city-level variables is straightforward: the relationship between the household-level regressors and demand is similar across China; so that when one province's observations are omitted, the resulting parameter estimate for the budgeting effect of, e.g., the gender of the household head accurately describes the relationship in the omitted province. The effects of the household-level factors included in these models—`age`, `educ`, `gender`, and `single`—are similar across provinces. In contrast, the same does not hold for population density, highway density, road vehicle stocks, and the other local condition and transport system indicators included from the CEIC source. In-

	clo	dur	ed	food	hou	med	other	trn
SH	-1.9	-0.5	-0.1	-1.1	0.0	1.8	0.4	1.0
CQ	-0.2	-0.0	-0.3	-1.5	0.0	-0.6	-1.2	-0.9
GD	-0.2	-0.1	0.1	-3.8	0.0	-0.0	-2.0	0.4
ZJ	-0.5	0.4	0.1	-2.9	0.1	-1.0	-0.7	-0.1
BJ	-0.6	0.1	-0.1	-2.1	0.2	-0.4	-1.2	0.3
HA	-2.5	0.3	-0.1	-1.8	-0.2	-0.4	-0.3	-0.3
HB	-2.0	0.1	0.0	-1.2	0.1	0.3	0.3	-0.3
SC	-2.4	-0.1	-0.1	-1.6	-0.1	-0.5	-0.4	-0.4
JS	-3.4	-0.1	-0.0	-2.1	0.0	0.1	-0.9	-0.7
YN	-3.2	0.1	0.0	-0.9	0.5	-0.4	-1.5	-0.4
AH	-2.2	-0.1	-0.0	-2.6	0.1	-0.8	-1.2	-0.2
LN	-1.1	-0.1	0.0	-1.8	0.2	-0.5	-2.0	-0.3
SX	-4.5	-0.1	-0.0	-2.2	0.4	-1.1	-0.6	0.1
GS	-2.4	-0.2	0.2	-2.4	-0.1	-0.2	-1.5	-0.3

	clo	dur	ed	food	hou	med	other	trn
SH	330.2	-7.8	35.0	156.3	21.9	-9.9	-4.6	-4.3
YN	-0.0	0.2	-1.1	11.5	-0.2	0.1	3.9	0.2
JS	104.7	3.2	16.3	4.7	5.1	-0.0	26.0	18.3
GD	13.7	-1.3	1.8	6.5	2.7	6.3	10.7	3.3
CQ	142.3	0.9	2.4	16.3	1.2	-2.0	-19.0	-2.4
ZJ	214.1	0.4	2.4	3.5	7.9	-0.2	24.2	-0.6
SX	-8.0	1.0	-2.2	-11.6	-0.2	-0.1	-2.0	-0.0
BJ	3.5	5.1	-1.2	16.0	14.9	17.6	12.6	-1.0
SC	-1.1	2.6	6.1	35.0	8.7	0.4	22.9	0.3
HB	62.5	7.7	14.3	8.0	0.8	5.7	43.7	0.2
GS	-2.3	0.1	-1.9	10.3	-0.5	-0.3	-2.1	-0.2
HA	67.2	0.3	16.2	2.1	7.4	-2.7	9.0	1.7
AH	153.8	5.2	19.9	3.8	0.2	5.4	40.3	-1.0
LN	-1.5	0.3	-0.3	3.1	1.8	-1.0	-1.4	-0.8

Figure 4-3: Percent difference in rmse for models `y4+hh` (top) and `y4+hh+city` (bottom), both relative to `y4`, geographical cross-validation.

stead, omitting a province from the training data leads to a biased parameter estimate: one that fits the mean effect across the included provinces, but poorly describes the relationship for households in the omitted province. The result implies that the influence of these factors differs from city to city, within the locations sampled by CHIP; in other words, the relationship between local conditions and households' transport expenditure is location-dependent. This also aligns with the failure, in the previous chapter (Section 3.6.4), to find significant parameter estimates for the effects of city-level variables such as `density` on $w^{\mathtt{trn}}$, even as F-tests showed that the variables were significantly related to overall budgeting. Another potential explanation is that the CEIC data are less proximate to individual CHIP households than the demographics drawn directly from CHIP responses—they describe conditions in the households' prefecture or county, rather than in their neighbourhood or block. However, as mentioned in Section 3.2.2, literature on the built environment has successfully measured effects for aspects of urban condition measured at this level of resolution.

This finding counsels strong caution in developing uses of EASI of the prescriptive type described in Section 4.2.3. A modeler who obtains data on, e.g., proximity of households to transit, might choose to include it in a model specification in order to test the demand effect of a hypothetical shift. But a significant parameter estimate in such a model may mislead, unless a validation check, such as the one described here, confirms the variable holds a consistent and not varying relationship to transport expenditures across and beyond the regions supplying the data available for estimation.

To go further, prediction errors for households in particular provinces are correlated with the magnitude of the city-level variables attached to those observations. Figure 4-4 on page 179, shows, for each city-level variable, the mean of its value across households in each province, versus the prediction error on those households when they are withheld from the training set. The AIDS model is compared to two

EASI specifications: one with, and one without province-level fixed effects. We see that, e.g., Shanghai (SH) households in the CHIP sample are located in districts with a mean highway density (`hwy_density`) among the highest in all of China, and much higher than the local highway density of other households in the survey data. These households' transport budget share are also predicted with the highest rmse if they are left out of the training set. Conversely, households surveyed in Liaoning (LN) and Gansu (GS) live, on average, in areas with low highway network density; and the prediction errors for their transport expenditure are relatively small. The Pearson's correlation coefficient for this relationship is $\rho = 0.39$ when controlling for province-level confounders, $\rho = 0.61$ when not—both lower than the analogous value of $\rho = 0.86$ for the AIDS model, where the variable is not included explicitly in the demand system.

So although the addition of city-level variables increases the sensitivity of prediction performance to the training data, the explicit inclusion of these measures of local condition in the EASI demand system absorbs some—again, only the mean effect across provinces—of their correlation with households' transport budgets. In addition to the heterogeneity of the influence of these variables (as discussed), this may suggest non-linearity of the relationships or interactions with other variables, income, or prices, that could be explored in future research.

4.4.3 Controlling for unobserved, province-level confounders

Finally, I highlight the occurrence in the EASI-CHIP models of a more general issue of importance when choosing model specifications for projection. When the goal is not projection but obtaining unbiased estimates of the influence of local conditions on household transport expenditure, there are standard requirements. Exogeneity is one, and so Chapter 3 discussed possible endogeneity of the available measures

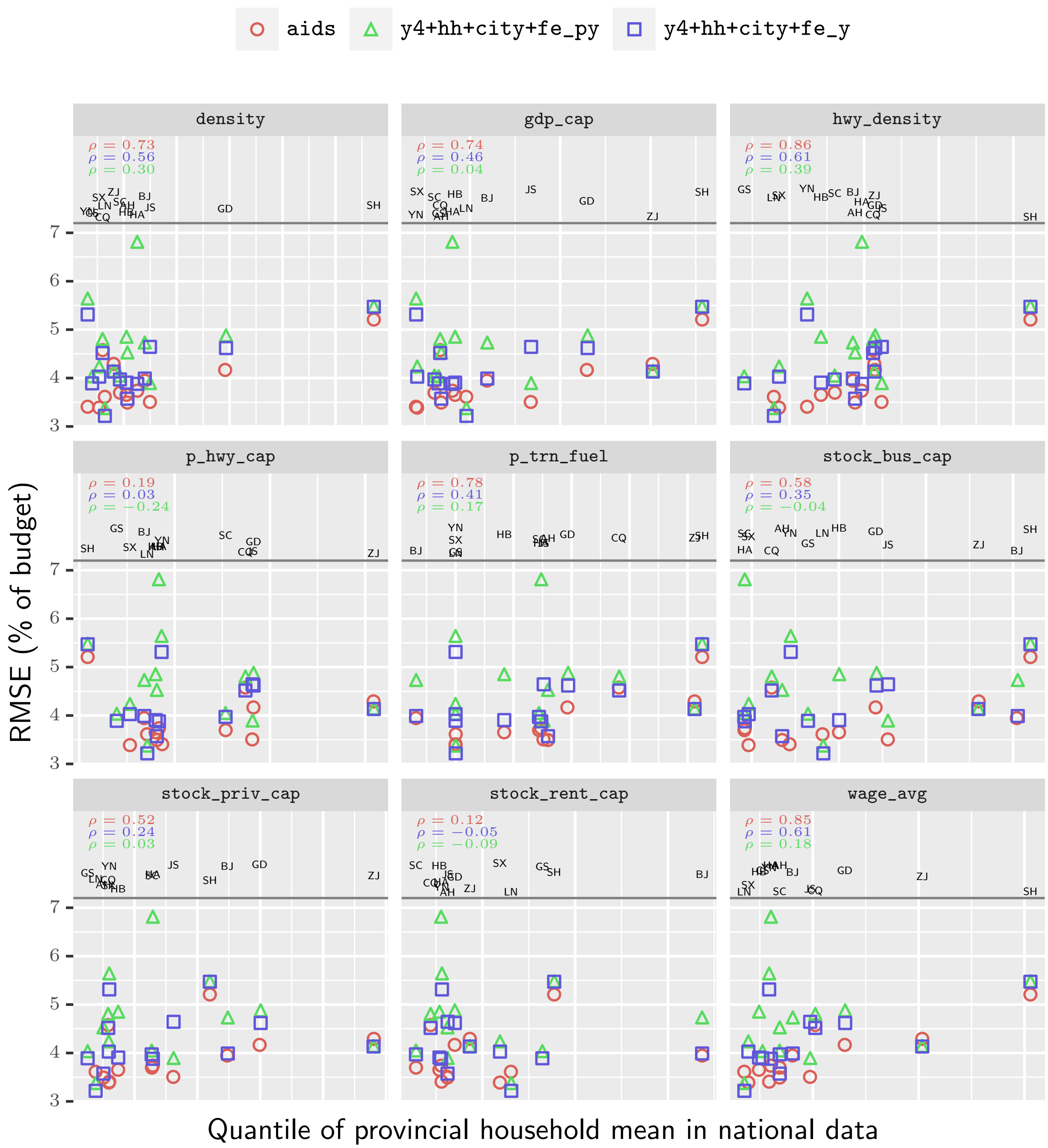


Figure 4-4: Correlation of rmse with city-level variables in AIDS and EASI models y4+hh+city with ("fe_py") and without ("fe_y") province fixed effects. Top portion of each panel gives the horizontal identifiers for markers at the same abscissa, and also the correlation coefficient, ρ, for each model.

Table 4.1: Fixed effect estimates for EASI models.

	y4	y4	y4+hh+city	y4+hh+city	y2+hh+city	y6+hh+city
Year F.E.	–	•	–	•	•	•
AH	−0.82	−0.44	6.12	1.80	1.69	1.81
BJ	0.00	0.00	0.00	0.00	0.00	0.00
CQ	−0.04	0.57	6.74	2.48	2.41	2.49
GD	−0.48	0.03	3.38	1.67	1.63	1.66
GS	−0.79	−0.88	4.80	1.05	0.97	1.06
HA	−0.56	−0.20	6.51	2.17	2.05	2.18
HB	−1.05	−0.79	5.85	1.53	1.41	1.53
JS	−1.42	−1.22	4.07	1.01	0.95	1.01
LN	−1.19	−0.51	4.37	0.96	0.87	0.96
SC	−1.27	−0.80	5.28	1.32	1.24	1.33
SH	1.16	2.28	−0.23	2.73	2.83	2.69
SX	−0.02	−0.92	5.55	1.24	1.11	1.26
YN	−0.76	−0.86	5.46	1.92	1.82	1.92
ZJ	−1.72	−0.13	1.82	0.89	0.89	0.88

and potential remedies (page 140). Another is control for omitted variable bias; so province and year dummies were included to capture the influence of unobserved attributes that differ from province to province but are time-invariant; and vice versa.

However, I find that this standard prescription (Bjerk 2009; Imai and Kim 2017) for accurate estimates of effect sizes negatively affects out-of-sample predictive performance. Just as adding city-level variables increases the sensitivity of predictions to the data used for estimation, adding province-level fixed effects to a model without them compounds the issue (Figure 4-5 on page 188); for nearly all consumption categories and validation segments, prediction errors grow, in some cases by over 100 %. Table 4.1 gives the coefficient estimates for the fixed effects, for several models. As noted, Beijing is chosen as a reference level. Other provinces have unobservable factors leading to shifts of 0.89 to 2.73 points of expenditure in households' budgets (column for `y4+hh+city`, with year F.E.). If the influence of unobserved factors is of a similar magnitude in other provinces, then the model can be expected to err by that amount; and unpredictably because the effect is, by construction, orthogonal to that of other variables.

There is no easy solution for modeling and projecting demand: without an understanding of which variables are unobserved, specifying a model without the effects will result in parameters biased in an unknown way. On the other hand, models with spatial fixed effects brings an error of a magnitude that, in the current data, is large in comparison to the transport budget share.

4.5 Discussion

This work has applied cross-validation methods to further explore the demand systems developed in Chapter 3. In addition to several extensions, to be discussed in Chapter 5, two categories of broader implications arise from the work: first, for modeling practice that incorporates economic systems of transport demand (Section 4.5.1), and second, for the collection of data to support better, valid characterizations of Chinese households' travel behaviour (Section 4.5.2).

4.5.1 Designing flexible demand systems for projection and assessment

In this analysis EASI demand systems, due to their ability to capture how consumption within categories responds to income in non-linear ways, offer a modest benefit over AIDS models of household demand. Higher-order functions of implicit utility (related to households' income), allows them to better capture how expenditure in the categories studied here varies across diverse populations. Due to their relative simplicity, I find that forms without demographic variables are suitable for use in out-of-sample projection or incorporation in predictive, aggregate models.

As a general prescription, modelers may use the methods provided here within any newly available data set, or with new demand formulations, to flag and identify problematic variables or aspects of variation across their sample that point to unobserved differences in the complex transport systems being modeled. For instance, if geographical cross-validation turns up high prediction error for a particular region, then other, out-of-sample regions with transport systems that can be considered similar in any sense may be watched for similar errors.

Choosing R. Chapter 3 identified third and fourth powers of implicit utility, y, as being significant to transport budget shares of the CHIP households. Omitting these powers from models biases coefficients, including those on the province-level dummies added to absorb the effects of unobserved factors—shown in Table 4.1 by the differences between the fixed effect for each province in model `y4+hh+city` compared to the restricted `y2+hh+city`. Adding additional (fifth and sixth) powers of y does not appear to change these parameters; the fixed effects for model `y6+hh+city` are much the same as for `y4+hh+city`. However, the $R = 6$ model in Section 4.4.1 more strongly outperforms the AIDS than one with $R = 4$.[5]

Household covariates. Because (per Section 4.4.2) the addition of household-level covariates such as the age, level of education or marital status of the household does significantly improve expenditure prediction errors, it appears feasible to employ EASI in settings where these are not available. For instance, in constructing synthetic populations for unsurveyed provinces, it is sufficient to simulate households income, x, from aggregate statistics and measures of inequality such as Gini indices. It is not additionally necessary to model household-level attributes, which would require difficult-to-obtain information on how these are correlated with income and one another. However, if this information is available, it may be used without raising

[5]Note that, unlike the model discussed in Section 3.6.1, the simply-specified EASI model in Section 4.4.1 does not include household-demographics, city-level variables, or province- or year fixed effects, and consequently has many fewer free parameters. This allows the y^5 and y^6 terms to contribute to better predictive performance in the latter model, while in the former they begin to exhaust the statistical power of the data.

concerns about prediction error.

Measures of local conditions and the built environment. Regarding the use of city-level covariates in out-of-sample predictions, the results raise caveats and warnings. EASI models including the particular set of variables used here performed worse in cross-validation than those without; worse also than AIDS models that did not include any local information. Moreover, the inclusion of provincial fixed effects, while helpful for unbiased parameter estimation, worsens the issue.

Modelers might choose to make educated guesses, equating out-of-sample regions with certain in-sample ones, assigning the same magnitude of influence from the unobservables. They may also explore and test local condition variables individually, and seek appropriate instruments to control for endogeneity. Another response, not pursued here, is to test interactions of the variables with one another, with household-level demographics, with income, and/or with prices—with the goal of identifying the specific interactions that capture the underlying heterogeneity in the way households' travel behaviour relates to their local context.

4.5.2 Implications for transport data collection

Questionnaire surveys such as CHIP, CGSS, or the China Residential Energy Consumption Survey (CRECS), remain expensive to conduct. CHIP, due to its division of the entire population into urban, rural, and rural-urban migrant households, missed the opportunity to surveys households in the latter two populations about expenditure by category. Use of other surveys that include both rural and urban households would strengthen the validity of conclusions about the nature of demand. Also, in designing a survey focused on the social conditions, family, and employment situations of households, and basic demographics of their members, the researchers of CHIP did not seek to ensure that the units of their stratified sampling process spanned the range of these transport- and planning-related variables, although Figure 4-4 suggests that they nevertheless covered the spectrum of certain variables fairly well.

This work identified that systematic differences across geographies—specifically provinces—places limits on the ability to draw empirical conclusions about the nature of demand (Chapter 3) and to project it in other contexts (this chapter). These were found at the provincial level, but may also or primarily manifest at other levels where transport system conditions and contextual factors vary, such as that of individual cities. At the province level, there were two types of differences, both of which can be addressed with improved data:

Unobserved attributes. The CEIC data source provides an aggregation of official statistics, which measure somewhat idiosyncratic concepts that do not well align with the concepts and measures from the literature on travel and the built environment (Section 3.2.2). Effort and resources could be applied to assembling or collecting data that align with these concepts, in a systematic way, and for locations matching the CHIP respondents. One potentially useful category of plentiful—if not conceptually aligned—data is referred to as "big" data—for instance, dispatch data from large logistics-matching platforms; satellite and remote-sensing data on vehicle positions and travel destinations; and information from smartphone-based services, including apps used to arrange shared mobility services. Exploration or selection of these sources can yield data for investigation of mobility patterns (Simini et al. 2012). These would, however, entail additional work in translation to the measures and concepts in the literature.

Heterogeneity in relationships. The former tasks are not necessarily for survey researchers—merely guided by their choice of sampling frames. Improving those frames, however, offers another way to provide improved data for demand characterization. Specifically, covering a small number of cities within each of a larger number of provinces would support investigation of how the influence of local conditions varies from place to place. Per S. Wang and J. Zhao (2018), discussed above, one strategy for household data collection would be to ensure that sampling frames spanned cities of different types, as identified from objective (data-based) classifications or clustering using aggregate information.

4.5.3 Conclusions

In developing implicit Marshallian demands and the exact affine Stone index demand system, Lewbel and Pendakur provided a means of characterising demand with more flexibility than earlier, still widely-used methods such as the AIDS. In Chapter 3, these features were used with household survey data to draw knowledge about budget shares and income elasticities from an household-level survey data set and public statistics.

As found here for the Chinese context, the new demand systems' flexibility confers both advantages and challenges in projection and assessment applications. The higher-order functions of income allow simply-specified EASI demands to be more robust to estimation on subsets of data, a valuable result in countries such as China where data without uniform nation-wide coverage must be used to understand transport demand. However, the flexibility to include a broad variety of covariates describing households and their local environment can lead to models that are fragile rather than robust. This issue has multiple sources: endogenous regressors, unobserved city- and province-level factors, and relationships between expenditure and context that vary in strength from province to province. The cross-validation methods used to reach these conclusions can also be used to test for such issues in other demand formulations and other data, supporting the development of valid models for standalone use and incorporation in modeling frameworks.

References

Bjerk, David (2009). "How Much Can We Trust Causal Interpretations of Fixed-Effects Estimators in the Context of Criminality?" In: *Journal of Quantitative Criminology* 25.4, pp. 391–417. ISSN: 1573-7799. DOI: 10.1007/s10940-009-9073-y (↪ p. 180).

Chen, Y.-H. Henry (2017). "The Calibration and Performance of a Non-homothetic CDE Demand System for CGE Models". In: *Journal of Global Economic Analysis* 2.1, pp. 166–214. DOI: 10.21642/JGEA.020103AF (↪ pp. 162, 163, 167, 196).

Cogneau, Denis and Anne-Sophie Robilliard (2007). "Growth, Distribution and Poverty in Madagascar. Learning from a Microsimulation Model in a General Equilibrium Framework". In: *Microsimulation as a Tool for the Evaluation of Public Policies. Methods and Applications*. Ed. by Amedeo Spadaro. Bilbao, ES: Fundación BBVA. chap. 3, pp. 73–111. ISBN: 978-84-96515-17-8 (↪ p. 163).

Deaton, Angus and John Muellbauer (1980). "An Almost Ideal Demand System". In: *The American Economic Review* 70.3, pp. 312–326. ISSN: 00028282. URL: http://www.jstor.org/stable/1805222 (↪ pp. 92, 167).

Dixon, Peter B., Robert B. Koopman, and Maureen T. Rimmer (2013). "The MONASH Style of Computable General Equilibrium Modeling: A Framework for Practical Policy Analysis". In: *Handbook of Computable General Equilibrium Modeling SET, Vols. 1A and 1B*. ed. by Peter B. Dixon and Dale W. Jorgenson. Vol. 1. Handbook of Computable General Equilibrium Modeling. Elsevier. Chap. 2, pp. 23–103. DOI: 10.1016/B978-0-444-59568-3.00002-X (↪ p. 167).

Fajgelbaum, Pablo D. and Amit K. Khandelwal (2016). "Measuring the Unequal Gains from Trade". In: *The Quarterly Journal of Economics* 131.3, pp. 1113–1180. DOI: 10.1093/qje/qjw013 (↪ p. 163).

Farber, Steven and Antonio Páez (2007). "A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations". In: *Journal of Geographical Systems* 9.4, pp. 371–396. ISSN: 1435-5949. DOI: 10.1007/s10109-007-0051-3 (↪ p. 169).

Henningsen, Arne (2011). *micEconAids: Demand Analysis with the Almost Ideal Demand System (AIDS)*. URL: https://cran.r-project.org/package=micEconAids (visited on 04/03/2018) (↪ p. 167).

Holz, Carsten A. (2004). "China's Statistical System in Transition: Challenges, Data Problems, and Institutional Innovations". In: *Review of Income and Wealth* 50.3, pp. 381–409. DOI: 10.1111/j.0034-6586.2004.00131.x (↪ pp. 47, 109, 165).

— (2013). "Chinese statistics: classification systems and data sources". In: *Eurasian Geography and Economics* 54.5–6, pp. 532–571. DOI: 10.1080/15387216.2014.908312 (↪ pp. 109, 165).

Imai, Kosuke and In Song Kim (2017). *When Should We Use Fixed Effects Regression Models for Causal Inference with Longitudinal Data?* Working paper. URL: http://web.mit.edu/insong/www/pdf/FEmatch.pdf (↪ p. 180).

Karplus, Valerie J., Sergey Paltsev, Mustafa H. Babiker, and John M. Reilly (2013). "Applying engineering and fleet detail to represent passenger vehicle transport in

a computable general equilibrium model". In: *Economic Modelling* 30, pp. 295–305. DOI: 10.1016/j.econmod.2012.08.019 (↪ pp. 46, 162, 196).
Kishimoto, Paul Natsuo (2012). "Passenger transport in China under climate constraints: general equilibrium analysis, uncertainty, and policy". S.M. thesis. Massachusetts Institute of Technology. HDL: 1721.1/78492 (↪ pp. 162, 164, 196).
Lewbel, Arthur and Krishna Pendakur (2009). "Tricks with Hicks: The EASI Demand System." In: *American Economic Review* 99.3, pp. 827–863. ISSN: 0002-8282. DOI: 10.1257/aer.99.3.827 (↪ pp. 92, 95, 126, 128, 160, 163, 185, 196).
Nijkamp, Peter, Aura Reggiani, and Wai Fai Tsang (2004). "Comparative modelling of interregional transport flows: Applications to multimodal European freight transport". In: *European Journal of Operational Research* 155.3. Traffic and Transportation Systems Analysis, pp. 584–602. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2003.08.007 (↪ p. 168).
Norwood, Bailey, Matthew C. Roberts, and Jayson L. Lusk (2004). "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions". In: *American Journal of Agricultural Economics* 86.4, pp. 1032–1043. DOI: 10.1111/j.0002-9092.2004.00651.x (↪ p. 169).
Paltsev, Sergey, Y.-H. Henry Chen, Valerie J. Karplus, Paul Natsuo Kishimoto, John M. Reilly, Andreas Loeschel, Kathrine von Graevenitz, and Simon Koesler (2016). "Reducing CO_2 from Cars in the European Union". In: *Transportation*. ISSN: 1572-9435. DOI: 10.1007/s11116-016-9741-3 (↪ p. 162).
Pradhan, Biswajeet (2010). "Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia". In: *Advances in Space Research* 45.10, pp. 1244–1256. ISSN: 0273-1177. DOI: 10.1016/j.asr.2010.01.006 (↪ p. 169).
Refaeilzadeh, Payam, Lei Tang, and Huan Liu (2009). "Cross-Validation". In: *Encyclopedia of Database Systems*. Ed. by Ling Liu and M. Tamer Özsu. Boston, MA: Springer US, pp. 532–538. ISBN: 978-0-387-39940-9. DOI: 10.1007/978-0-387-39940-9_565 (↪ p. 168).
Rengaraju, V. R. and V. Thamizh Arasan (1992). "Modeling for Air Travel Demand". In: *Journal of Transportation Engineering* 118.3, pp. 371–380. DOI: 10.1061/(ASCE)0733-947X(1992)118:3(371) (↪ p. 168).
Rimmer, Maureen T. and Alan A. Powell (1996). "An implicitly additive demand system". In: *Applied Economics* 28.12, pp. 1613–1622. DOI: 10.1080/000368496327589 (↪ pp. 92, 167).
Simini, Filippo, Marta C. González, Amos Maritan, and Albert-László Barabási (2012). "A universal model for mobility and migration patterns". In: *Nature* 484, p. 96. DOI: 10.1038/nature10856 (↪ p. 184).
Stagars, Manuel (2016). "Tools to Standardize and Monitor Statistical Capacity and Data Quality". In: *Data Quality in Southeast Asia: Analysis of Official Statistics and Their Institutional Framework as a Basis for Capacity Building and Policy Making in the ASEAN*. London: Palgrave Macmillan UK, pp. 7–20. ISBN: 978-1-137-60063-9. DOI: 10.1057/978-1-137-60063-9_2 (↪ p. 165).
Tovar Reaños, Miguel A. and Nikolas M. Wölfing (2018). "Household energy prices and inequality: Evidence from German microdata based on the EASI demand

system". In: *Energy Economics* 70, pp. 84–97. ISSN: 0140-9883. DOI: 10.1016/j.eneco.2017.12.002 (↪ p. 163).

Wang, Shenhao and Jinhua Zhao (2018). "Trajectories of Urban Development & Motorization: Clustering 287 Chinese Cities". Manuscript (in preparation) (↪ pp. 22, 164, 185).

Wenger, Seth J. and Julian D. Olden (2012). "Assessing transferability of ecological models: an underappreciated aspect of statistical validation". In: *Methods in Ecology and Evolution* 3.2, pp. 260–267. DOI: 10.1111/j.2041-210X.2011.00170.x (↪ p. 169).

Wood, Benjamin D.K., Carl H. Nelson, and Lia Nogueira (2012). "Poverty effects of food price escalation: The importance of substitution effects in Mexican households". In: *Food Policy* 37.1, pp. 77–85. ISSN: 0306-9192. DOI: 10.1016/j.foodpol.2011.11.005 (↪ p. 163).

Zhang, Ming, Hailin Mu, Gang Li, and Yadong Ning (2009). "Forecasting the transport energy demand based on PLSR method in China". In: *Energy* 34.9, pp. 1396–1400. ISSN: 0360-5442. DOI: 10.1016/j.energy.2009.06.032 (↪ p. 168).

4.A Figure

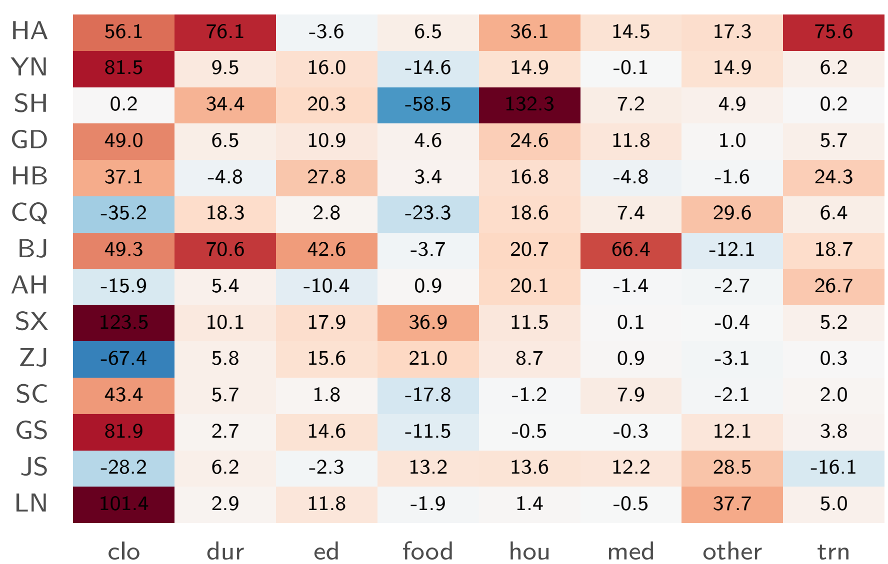


Figure 4-5: Percent difference in rmse for model `y4+hh+city`, comparing a model with province and year fixed effects to one without province fixed effects.

Chapter 5

Conclusion

A systems view of transportation emphasizes differences in context that will lead China's transport system evolution to proceed along a course distinct from the historical experience of wealthy, developed countries. Specifically, the combined influence of economic growth, inequality, urban form, culture, technology options, and institutions will lead to distinct patterns in transport activity growth, its environmental and other impacts, and the responses that emerge to manage these impacts. To deepen our understanding of these complex systems will require methods and analytical insights that recognize that the past is not always prologue.

In order to improve understanding of this complex system, this thesis has provided improved methods to characterize and project transport demand and its variation across cities, provinces, and modes; and to analyze policies that target the environmental impacts of growth. To conclude, I review the research in relationship to the framing established in Chapter 1 (Section 5.1); highlight the contributions to research literatures and modeling practice (Section 5.2); comment on the policy implications of the findings (Section 5.3); and detail some concrete studies (Section 5.4) and a broader agenda (Section 5.5) of research arising from this work.

5.1 Summary of the research and key insights

In Chapter 2, I developed methods for modeling road transport energy use and pollutant emissions in multiple sub-sectors and at the provincial level in China, within an economy-wide computable general equilibrium (CGE) model, itself embedded in an integrated assessment framework. This enabled comparison of the effects of emissions standards policies (focused on technology in road vehicles that are sold, used, and scrapped differently across provinces) and carbon pricing (which affects all CO_2-emitting energy use across sectors of the economy, including transport). I found that the instruments are complementary: carbon pricing yielding a co-benefit of economy-wide emissions reductions, but less in transport where CO_2 mitigation costs are higher; and emissions standards highly effective, but only for the transport share of emissions, which remains smaller in many Chinese provinces than in other, international contexts.

In Chapter 3, I looked into the nature of transport demand at the household level in China, by giving a new application of recently-developed, flexible, Exact affine Stone index (EASI) demand systems, wherein the expenditure share for transport is estimated alongside the shares for other categories of consumption. Using household data from a survey with national coverage, I produced new estimates of the Engel curve of transport expenditure and of the income elasticity of transport demand. Crucially, I found that the flexible demand system was able to capture a high-order empirical relationship between rising income (total expenditure) and households' transport demands; one that is not possible in older, linear forms. Also, budget shares are significantly impacted by local economic conditions, the built environment, and the characteristics of transport systems, which can be incorporated in detail using the new formulation. I demonstrated that, across certain groups of households, differences in income, via the income elasticity of demand, explain only a small part of the variation in transport budget shares, while the remainder is associated with local conditions, both observed and unobserved; with prices; and with year-specific unobserved factors.

Chapter 4 further examined the performance of flexible demand systems. By training models with household observations withheld by provinces, then testing the prediction performance of the models, I found that flexible demand systems can offer consistent, though modest improvements in out-of-sample performance compared to the widely-used Almost Ideal demand system (AIDS). Including local urban and transport system characteristics in specifications, while offering an explicit representation of demand-relevant concepts, also makes the out-of-sample performance of models strongly sensitive to coverage of the data used for estimation.

Chapter 1 introduced overarching questions of resolution and scope in modeling of transport systems: what insights and analytical capacity are gained by moving from national aggregates to greater detail; and what level of resolution is necessary in order to study the impacts of important contextual aspects—policy instruments, or geographical heterogeneity in transport systems?

The thesis provides insights toward both questions. First, by increasing detail in the transport sector of a CGE framework with national resolution, Chapter 2 showed the province-to-province variation in the contribution of different transport modes to emissions mitigation under overlapping policies (which also vary at the provincial level). In Chapter 3, evidence from the household microdata pointed to a lower, yet wider range for the share of transport in household budgets, and smaller income elasticities, compared to prior work that has relied on province-level data for a small number of income groups. Chapter 4 showed that new methods with increased flexibility in this income-demand relationship led to reduced prediction errors when used outside the contexts supplying the data for model estimation.

Across both parts, the primary added difficulty was in obtaining broad-coverage, high-resolution data on the transport activity, impact, and city condition measures. In

Chapter 2, this involved detecting and correcting for anomalies arising from statistical agencies' treatment of transport time series. In Chapters 3 and 4, the available data were an imperfect match to the concepts identified by previous travel behavior research, and the use of a social science survey limited analysis to a category of total transport expenditure that did not resolve spending on vehicles, fuels, and other specific modes.

On the second question, Chapter 2 provides projections of future air pollutant emissions that reflect significant provincial variations. In integrated assessment of the health and economic impacts of transport policies—that are, in the Chinese context, also set by local governments—this approach improves on emissions rates projected as nationwide averages. Chapter 3 showed that improved flexibility in demand system models is required to capture actual complexity in the influence of income growth on Chinese households' budgeting decisions. Finally, Chapter 4 provided evidence that influence of the built environment and other local factors on households' transport budgets differs in strength from place to place, necessitating models and data that can describe this variation and carry it forward into projection.

5.2 Contributions to research and practice

CGE models such as Economic Projection and Policy Analysis (EPPA) model and the China Regional Energy Model (C-REM) remain an important tool for assessing transport policy in the context of broader economic activity and economy-wide climate policy. Although developed for the China-provincial C-REM, the transport-sector disaggregation, and associated calibration procedures, and linkage through energy-basis emissions factors developed in Chapter 2 are portable to other CGE models. The need to separate freight and passenger road transport, in particular, is acute as new technologies such as mobility sharing and automated vehicles will have starkly different impacts on these subsectors.

By measuring a different dependent concept, Chapter 3 provides a thought-provoking contrast to transport economics literature that has largely produced demand elasticity estimates focused on gasoline or vehicle-distance travelled (VDT) and—especially in China—operated at the national aggregate level. Households' total transport expenditure varies with income in different ways, and with a greater elasticity, than these other sub-quantities; the work motivates further attention to variation in the composition of transport budgets. While this basket of varied goods and services presents challenges to analysis (as discussed in Section 3.7.1), it may also present an opportunity. Research on trajectories of passenger-distance travelled (PDT) or vehicle ownership per capita, has drawn on data from specific historical and national transport systems. Changes in context threaten these regularities: not only the rapid change in a country such as China, but new technologies and business models. For instance, the advent and spread of ride-hailing services are already noted to affect personal vehicle ownership; in the longer term these, or technologies such as shared, autonomous vehicles, may permanently separate China and other less-developed countries from the historical, Western path of motorization. On the other hand, individuals facing these new transport options—each with its supply and price characteristics—will continue to need to divide a fixed money budget between travel and other needs and wants, including housing, food, education, and medical care. This raises the possibility that transport expenditure may be a measure with more stability than other measures tied to historical transport mode options.

The chapter also provides a concrete example that, in countries such as China that yet lack large-scale, sophisticated nationwide travel surveys, unconventional sources such as other survey data (CHIP) can be exploited for insight into transport behaviour—even though they were not originally collected with any focus on transportation.

The EASI models revealed that income variation only explains a fraction of variation in transport spending. While the data available for the current work were imperfect measures of local conditions, and significant parameter estimates were not obtained, the models echoed the existence of links between these conditions and households' transport budgets, which are associated with a large portion of budget share differences. This work begins to connect literatures on travel behavior—in which nuanced concepts of the built environment are linked to discrete measures of travel behavior—with aggregate modeling literature that has balanced economic or other representations of transport demand with other categories of consumption that together comprise all final demand.

Finally, Chapter 4 concluded with specific directions for collection of data on transport system conditions and individuals' transport activity (including expenditure) that would support more precise estimates of the influence of local conditions. For modelers seeking to adopt EASI as a replacement for AIDS in standalone and framework applications, the chapter provided a prescription for designing specifications that exploit the promised flexibility of the newer demand system, and at the same time a framework for validating those specifications against very real threats to external validity.

5.3 Policy implications

Beyond the direct conclusions about the CO_2 pricing and emissions standards policy instruments compared in Chapter 2, Chapters 3 and 4 suggests some implications, if not direct prescriptions, for transport policymaking in China and other countries.

Pendakur (2009) notes that "in typical consumer demand models, observables like prices, expenditure and household demographics explain no more than half the variation in budget shares." In a similar way, it was shown in Section 3.6.3 that a large portion of the remainder of transport budget share variation can be associated with observable city-level attributes, and much of the rest (up to two percentage points of total expenditure, compared to a transport budget share that reaches a maximum of 7.5 points) with unobserved province- and possibly city-level attributes for which data might be collected.

This result reflects the diversity of transport system conditions across China's cities and provinces. It suggests that there is broad scope for city governments to affect the course of transport activity growth and transport system evolution in the future, by ensuring urban development guides households' travel expenditure in desirable ways. While specific impacts of travel demand are below the resolution of the data in this paper, the result motivates closer attention to these links and related instruments for mitigating the impacts of demand. At the same time, the high income elasticity of demand cautions that gross domestic product (GDP) growth in the coming years will bring an even greater rise in households' transport spending that, unmanaged, could worsen already severe issues of congestion, pollution, land use, and others.

The second implication is that these channels of influence may vary significantly from city to Chinese city. Again, causal estimates are beyond the scope of this work; but the results of Chapter 4 suggest that, for instance, households in a dense, centrally-administered city such as Shanghai, and cities in a poorer province such as Yunnan, change their travel behavior in distinct ways in response to greater population density. This means that different types of cities may find distinct policy levers to be most effective in addressing the consequences of growing transport demand. While existing research (J. Zhao and Z. Wang 2014) finds that emulation/variation (as discussed for vehicle license plate (VLP) policies in Section 1.3) and central-government directives both play roles in transferring policy ideas from city to city, both of these approaches risk leading cities to adopt measures that, while successful elsewhere, may do little to influence the transport behavior of local households.

5.4 Direct extensions to the thesis

Considering the endpoints of the three chapters together, some direct ideas for research emerge. I give two examples here to illustrate how such extensions could further enrich the bodies of literature that motivated the current work.

Improving CGE policy realism with EASI demands. The investigation of household-level demand in Chapter 3 was motivated, in part, by questions about the realism of parameterizations of aggregate demand by representative agents in each region (country or province) of CGE models, such as the C-REM employed in Chapter 2, and the EPPA model used in earlier work such as Kishimoto (2012). The EASI demand system was derived (Lewbel and Pendakur 2009) so as to be consistent under aggregation, meaning that an EASI demand formulation could substitute for a simpler representation of final demand in a general-equilibrium model without compromising the consistency of the model's economic logic. Since I found in Chapter 3 that Chinese households' transport demand is indeed a higher-order function of income, the question arises of how this relationship affects the cost and impact of climate or energy policies.

Existing frameworks for CGE modeling such as GAMS/MPSGE (Rutherford 1999) allow only constant elasticity of substitution (CES) as a built-in specifications; non-homothetic formulations require manual implementation (Chen 2017; Karplus et al. 2013), or the construction of new models (Caron et al. 2017). Either of these approaches could be taken to implement the Chapter 3 models or similar specifications in a province-level, general equilibrium model of China. Scenario simulations in both this and a less-flexible (homothetic) base model would then reveal how a more realistic description of household demand alters policy impacts and their interaction with income growth.

Characterizing variation in the expenditure—built environment link. One finding of Chapter 4 was that the exclusion of household observations from China Household Income Project (CHIP) provinces worsened the out-of-sample predictive performance of the resulting models. In Section 4.4.2 and Figure 4-4, this performance degradation was shown to be correlated to the city-level regressors included in the EASI models, across the CHIP provinces. A more detailed comparison of this type would have immediate practical value to stakeholders seeking to collect survey data or conduct targeted studies in order to better characterize China's transport system.

In particular, the root mean squared error (rmse) metric (or others) could be modeled by regression on observable attributes of provinces, cities, or households. The resulting parameters, applied to observable data for non-CHIP regions, would yield a first-order prediction of how poorly demand systems might predict transport expenditure in those areas—in short, how 'unusual the transport systems and individual transport behavior in such regions. This would allow future research and data collection to selectively undersample 'unremarkable' areas, while devoting resources to exploring areas where transport may be related to local context in idiosyncratic ways.

5.5 Future work

Within the transport sector, efforts to advance sustainability must contend with the rapid growth and motorization of demand in the emerging economies of low- and middle-income countries. M. Burke et al. (2016) describe current areas for advance in climate change economics, including the prospects and impacts for mitigation options in the developing world; distributional effects on lowest-income households whose quality of life must be improved to make development sustainable; and policies beyond the first-best instruments on which research has historically focused.

This thesis developed methods for using existing data sources—especially household microdata from social surveys—to characterize the drivers of transport demand in China, with the goal of informing models used in policy and planning decisions. While the work has focused on China in particular, the calls of M. Burke et al. and others motivate a focus on low- and middle-income countries where similar factors challenge transport research: rapid economic growth, early-stage motorization, evolving transport systems, data quality limitations, and idiosyncratic policy instruments and processes. Limitations of the individual essays were addressed in Sections 2.6.3, 3.7.2 and 4.5. Considering how some of these might be relaxed reveals three themes of potential future research:

1. What do existing data and relationships suggest about patterns of mobility that are likely to emerge in low- and middle-income countries as they evolve in the 21st century? How do projections compare across models, and what can be gained from adopting more empirically-driven model settings?
2. To what extent could new technologies and new policy instruments, interacting with heterogeneity in behaviours and preferences, influence the course of transport motorization in these countries?
3. How do #1 and #2 contribute to uncertainty in projections of transport activity, in particular in integrated assessment models that link transport activity to trends in the broader economy?

To make progress on these questions requires the use of detailed and disaggregate data that are not typically linked to large-scale integrated assessment models (IAMs)—which often represent activity at the level of national or regional totals, and are calibrated using aggregate data. In order to employ and connect these data, future research can investigate methods, including from the econometrics of consumer demand and policy analysis, that may not have been applied to energy or transport research; or enhance traditional demand modeling methods to allow them to act on unconventional data and in new contexts. Chapters 3 and 4 are examples of such research.

This detailed work can add conceptual richness and new insights to existing knowledge about the economic growth–transport demand relationship (#1), and can yield tools for analyzing and comparing scenarios of policy and technology adoption in motorizing countries or their regions or cities (#2). The empirical foundation and identified dynamics would then support efforts to improve IAM representation of the same impacts (#3)—through changes to functional relationships, calibration methods, and/or data.

5.5.1 Household-level modeling of transport demand systems

Several extensions to the thesis research can address themes #1 and #2. Besides CHIP, other survey efforts such as the China Household Finance Survey (CHFS) (Gan et al. 2014) and China General Social Survey (CGSS) (Bian and L. Li 2012) have included questions about household expenditure in their questionnaires—respectively, in 2011, and annually since 2003. The methods of Chapters 3 and 4 are applicable to such data sets, to the extent they contain a sufficient number of subcategories of expenditure to offer a basis for comparison.[1] Comparison of models estimated on different survey datasets, such as CHIP versus CGSS, will create richer knowledge about the stability or variation of demand across time periods and geographies. Beyond China existing, similar consumption data such as National Sample Survey (India) (NSS) and General Household Survey (Nigeria) (GHS), supports research to estimate analogous models and compare across a diversity of lower-income countries. Within countries,

[1]Concerns about the social and political consequence of county-specific information have led some non-government survey groups to restrict the location data used here to match urban variables (see e.g. http://www.chinagss.org/index.php?r=index/artabout&aid=18), which may limit this approach.

Second, by breaking the transport component of household expenditure—here treated as a single category—into several goods and services, demand system modeling can allow investigation of shifts in household travel patterns across income distributions. Finally, with wide-coverage databases of policy instruments assembled from primary source documents, demand system regressors can be expanded to introduce explicit policy variables. This will allow direct identification of the effects of road pricing, VLP restrictions, and other instruments seen in recent history; across a broad set of urban contexts within and across countries.

5.5.2 Increasing transport realism in IAMs

There is great potential to improve the realism of models by carefully and systematically estimating relationships of interest and incorporating them into model structure and/or parameterization. McCollum et al. (2017), *inter alia*, emphasize that realism in consumer behaviour—*ie.*, fidelity to sub-model scale empirical patterns—in IAMs can alter the projected timing and path of energy and technology transitions. Past work with CGE models—the C-REM with provincial detail, as in Chapter 2, and the EPPA model (Kishimoto, Paltsev, et al. 2012)—emphasizes that the quantity and prices of inputs to transport services are endogenous with other flows across the entire economy. However, Yeh et al. (2016) and others describe the considerable variation in the functional relationships encoded in global models of transport, energy and greenhouse gas (GHG) emissions. For instance, in the International Institute for Applied Systems Analysis (IIASA) Model for Energy Supply Systems and their General Environmental impact (MESSAGE) framework a highly-aggregate CGE model is used to project overall economic growth, which is then apportioned to specific technologies within sectors by linear-programming optimization.

Theme #3 points at the importance of understanding how these methodological variations affect models' ability to capture aspects of consumer behaviour, heterogeneity, and technology and policy responses. Focused, multi-model comparisons can both yield insights for consumers of model-based knowledge and assessment—who must respond to disagreement about projections and policy effects. Modellers working in different frameworks also require distinct methods to identify whether emergent, system-level attributes in their simulations accord or conflict with empirical findings, and tailored yet consistent methods for bringing projections in line with such findings.

5.5.3 Simulated populations and synthetic controls

Separate from their incorporation in integrated assessment models, econometric models including consumer demand systems such as EASI can support policy analysis through the method of synthetic controls (Abadie et al. 2010). In this approach, a counterfactual is simulated by combination of available data for untreated regions. This method holds promise for analysis of novel transport policy instruments; namely the VLP restrictions in some large Chinese cities.

In order to study these instruments, the work described in this thesis can be extended by adopting the CGSS data set the period up to 2015, such that households treated with VLP restrictions enter the sample (see Table 1.2 on page 30). Contrasting these households against synthetic controls would produce new estimates of the impact of VLP policies on household transport demand. As the policy has been floated for consideration in other large cities in the developing world, this analysis would speak to the question (within theme #2) of how the VLP policy instrument can support transitions to sustainable transport.

Uncertainty in model-based projections of transport activity growth in lower-income countries, and its response to policy, poses challenges for the assessment of pathways to be followed to achieve both the Sustainable Development Goals and countries' nationally-determined contributionss (NDCs) towards the Paris Agreement under the United Nations Framework Convention on Climate Change (UN FCCC). In this thesis, I developed empirical methods that can explit unconventional data to derive local insights into the nature of transport demand in China's cities; as well as modeling techniques for enhancing spatial and transport sector resolution in CGE models, connecting micro-scale decisions and sub-national policies to aggregate, macro-scale impacts and outcomes. The work points the way to increased use of novel data sources to better characterize transport system evolution in the countries that will be responsible for the bulk of growth in activity and environmental impacts throughout the 21st century, and to analyze policies that will address these impacts.

Bibliography

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program". In: *Journal of the American Statistical Association* 105.490, pp. 493–505. DOI: 10.1198/jasa.2009.ap08746 (↪ p. 201).

Adcock, Robert and David Collier (2001). "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research". In: *American Political Science Association* 95.3 (03), pp. 529–546. ISSN: 1537-5943. DOI: 10.1017/S0003055401003100 (↪ p. 21).

Bass, Frank M. (1969). "A New Product Growth for Model Consumer Durables". In: *Management Science* 15.5, pp. 215–227. DOI: 10.1287/mnsc.15.5.215 (↪ p. 24).

Bian, Yanjie and Lulu Li (2012). "The Chinese General Social Survey (2003-8)". In: *Chinese Sociological Review* 45.1, pp. 70–97. DOI: 10.2753/CSA2162-0555450104 (↪ p. 199).

Bonus, Holger (1973). "Quasi-Engel Curves, Diffusion, and the Ownership of Major Consumer Durables". In: *Journal of Political Economy* 81.3, pp. 655–677. ISSN: 0022-3808, 1537-534X. URL: http://www.jstor.org/stable/1831030 (↪ p. 24).

Burke, M., M. Craxton, C. D. Kolstad, C. Onda, H. Allcott, E. Baker, L. Barrage, R. Carson, K. Gillingham, J. Graff-Zivin, M. Greenstone, S. Hallegatte, W. M. Hanemann, G. Heal, S. Hsiang, B. Jones, D. L. Kelly, R. Kopp, M. Kotchen, R. Mendelsohn, K. Meng, G. Metcalf, J. Moreno-Cruz, R. Pindyck, S. Rose, I. Rudik, J. Stock, and R. S. J. Tol (2016). "Opportunities for advances in climate change economics". In: *Science* 352.6283, pp. 292–293. ISSN: 0036-8075. DOI: 10.1126/science.aad9634 (↪ pp. 197, 198).

Caron, Justin, Valerie J. Karplus, and Giacomo A. Schwarz (2017). *Modeling the Income Dependence of Household Energy Consumption and its Implication for Climate Policy in China.* JP Report 314. Cambridge, MA, US: MIT Joint Program on the Science and Policy of Global Change. HDL: 1721.1/111814 (↪ pp. 94, 120, 196).

Chen, Y.-H. Henry (2017). "The Calibration and Performance of a Non-homothetic CDE Demand System for CGE Models". In: *Journal of Global Economic Analysis* 2.1, pp. 166–214. DOI: 10.21642/JGEA.020103AF (↪ pp. 162, 163, 167, 196).

Chen, Y.-H. Henry, Sergey Paltsev, John M. Reilly, Jennifer F. Morris, and Mustafa H. Babiker (2015). *The MIT EPPA 6 Model: Economic Growth, Energy Use, and Food Consumption.* Tech. rep. 278. Cambridge, MA, US: MIT Joint Program on the Science and Policy of Global Change. HDL: 1721.1/95765 (↪ pp. 22, 91).

Cheng, Yang (2013). *Quota details of Tianjin's car-buying restrictions released.* Xinhuanet. URL: http://news.xinhuanet.com/english/china/2013-12/22/c_132987003.htm (visited on 12/18/2015) (↪ p. 30).

China Daily (2014). *Cities put brakes on new car license plates.* China Daily. URL: http://www.chinadaily.com.cn/m/guizhou/guiyang/2014-01/08/content_17220396.htm (visited on 12/18/2015) (↪ p. 30).

Creutzig, Felix, Patrick Jochem, Oreane Y. Edelenbosch, Linus Mattauch, Detlef P. van Vuuren, David McCollum, and Jan Minx (2015). "Transport: A roadblock to climate change mitigation?" In: *Science* 350.6263, pp. 911–912. ISSN: 0036-8075. DOI: 10.1126/science.aac8033 (↪ p. 18).

Dargay, Joyce, Dermot Gately, and Martin Sommer (2007). "Vehicle ownership and income growth, worldwide: 1960-2030". In: *The Energy Journal* 28.4, pp. 143–170. DOI: 10.5547/ISSN0195-6574-EJ-Vol28-No4-7 (↪ pp. 24, 25, 112, 113).

Feng, Suwei and Qiang Li (2013). "Car Ownership Control in Chinese Mega Cities: Shanghai, Beijing and Guangzhou". In: *Journeys.* Ed. by George Sun, Evan Gwee, Foo Jong Ai, and Augustine Low. 10. Singapore, SG: LTA Academy, pp. 40–49. URL: http://www.lta.gov.sg/ltaacademy/LA-I10Journeys_Sep13.htm (↪ p. 30).

Fishman, Elliot (2016). "Bikeshare: A Review of Recent Literature". In: *Transport Reviews* 36.1, pp. 92–113. DOI: 10.1080/01441647.2015.1033036 (↪ p. 17).

Gan, Li, Zhichao Yin, Nan Jia, Shu Xu, Shuang Ma, and Lu Zheng (2014). *Research Report of China Household Finance Survey 2012.* Berlin, DE: Springer. ISBN: 978-3-642-38151-5. DOI: 10.1007/978-3-642-38151-5 (↪ p. 199).

Gärling, Tommy and Linda Steg, eds. (2007). *Threats From Car Traffic To The Quality of Urban Life. Problems, Causes, and Solutions.* Oxford, UK: Elsevier. ISBN: 978-0-08-044853-4 (↪ p. 19).

Golder, Peter N and Gerard J Tellis (1998). "Beyond diffusion: An affordability model of the growth of new consumer durables". In: *Journal of Forecasting* 17, pp. 259–280 (↪ p. 24).

Gota, Sudhir, Cornie Huizenga, Karl Peet, and Gavin Kaar (2016). *Nationally-Determined Contributions (NDCs) Offer Opportunities for Ambitious Action on Transport and Climate Change.* Tech. rep. Shanghai, CN: Partnership on Sustainable Low Carbon Transport. URL: http://www.ppmc-transport.org/overview_indcs/ (↪ p. 18).

Haagen-Smit, A. J. (1952). "Chemistry and Physiology of Los Angeles Smog". In: *Industrial & Engineering Chemistry* 44.6, pp. 1342–1346. DOI: 10.1021/ie50510a045 (↪ p. 18).

Ho, Mun S. and Chris P. Nielsen, eds. (2007). *Clearing The Air: The Health and Economic Damages of Air Pollution in China.* Cambridge, MA, US: MIT Press. ISBN: 9780262083584 (↪ p. 28).

Hoareau, Stéphane, Guy Lacroix, Mirella Hoareau, and Luca Tiberti (2012). *Ex-*

act Affine Stone Index Demand System in R: The EASI package. URL: https://github.com/cran/easi/blob/master/inst/doc/easi.pdf (↪ pp. 115, 117, 126, 230).

Hou, Quan and Si-Ming Li (2011). "Transport infrastructure development and changing spatial accessibility in the Greater Pearl River Delta, China, 1990–2020". In: *Journal of Transport Geography* 19.6. Special section on Alternative Travel futures, pp. 1350–1360. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2011.07.003 (↪ pp. 28, 102).

Hsieh, I-Yun (Lisa), Paul Natsuo Kishimoto, and William Green (2018). "Incorporating multiple uncertainties into projections of Chinese private car sales & stock". In: *Transportation Research Record. Journal of the Transportation Research Board*. In press. ISSN: 0361-1981 (↪ pp. 26, 27).

Huo, Hong and Michael Wang (2012). "Modeling future vehicle sales and stock in China". In: *Energy Policy* 43, pp. 17–29. DOI: 10.1016/j.enpol.2011.09.063 (↪ p. 24).

Karplus, Valerie J., Sergey Paltsev, Mustafa H. Babiker, and John M. Reilly (2013). "Applying engineering and fleet detail to represent passenger vehicle transport in a computable general equilibrium model". In: *Economic Modelling* 30, pp. 295–305. DOI: 10.1016/j.econmod.2012.08.019 (↪ pp. 46, 162, 196).

Kishimoto, Paul Natsuo (2012). "Passenger transport in China under climate constraints: general equilibrium analysis, uncertainty, and policy". S.M. thesis. Massachusetts Institute of Technology. HDL: 1721.1/78492 (↪ pp. 162, 164, 196).

Kishimoto, Paul Natsuo, Sergey Paltsev, and Valerie J. Karplus (2012). *The Future Energy and GHG Emissions Impact of Alternative Personal Transportation Pathways in China*. JP Report 231. Cambridge, MA, US: MIT Joint Program on the Science & Policy of Global Change. HDL: 1721.1/73607 (↪ pp. 46, 200).

Knittel, Christopher R. (2012). "Reducing Petroleum Consumption from Transportation". In: *Journal of Economic Perspectives* 26.1, pp. 93–118. DOI: 10.1257/jep.26.1.93 (↪ p. 18).

Lewbel, Arthur and Krishna Pendakur (2009). "Tricks with Hicks: The EASI Demand System." In: *American Economic Review* 99.3, pp. 827–863. ISSN: 0002-8282. DOI: 10.1257/aer.99.3.827 (↪ pp. 92, 95, 126, 128, 160, 163, 185, 196).

Li, Fusheng (2014). *Cities put brakes on new car license plates*. China Daily. URL: http://europe.chinadaily.com.cn/business/2014-01/06/content_17217743_2.htm (visited on 12/18/2015) (↪ p. 30).

McCollum, David L., Charlie Wilson, Hazel Pettifor, Kalai Ramea, Volker Krey, Keywan Riahi, Christoph Bertram, Zhenhong Lin, Oreane Y. Edelenbosch, and Sei Fujisawa (2017). "Improving the behavioral realism of global integrated assessment models: An application to consumers' vehicle choices". In: *Transportation Research Part D. Transport and Environment* 55, pp. 322–342. ISSN: 1361-9209. DOI: 10.1016/j.trd.2016.04.003 (↪ pp. 22, 200).

Molina, Luisa T. and Mario J. Molina, eds. (2002). Vol. 2. Alliance for Global Sustainabilty book series. Dordrecht, The Netherlands: Kluwer. ISBN: 978-1-4020-0452-0. DOI: 10.1007/978-94-010-0454-1 (↪ p. 19).

Mostashari, Ali and Joseph M. Sussman (2009). "A Framework for Analysis, Design and Management of Complex Large-Scale Interconnected Open Sociotechnological Systems". In: *International Journal of Decision Support System Technology* 1.2, pp. 53–68. DOI: 10.4018/jdsst.2009040104 (↪ p. 17).

Mu, Xuequan (2011). *Car license plate limit triggers controversy in SW China city*. Xinhuanet. URL: http://news.xinhuanet.com/english2010/china/2011-07/18/c_13991045.htm (visited on 12/18/2015) (↪ p. 30).

National Bureau of Statistics of China (2012). *China Statistical Yearbook 2011*. Beijing, CN. URL: http://www.stats.gov.cn/tjsj/ndsj/2011/indexeh.htm (↪ p. 27).

Nielsen, Chris P. and Mun S. Ho, eds. (2013). *Clearer Skies Over China. Reconciling Air Quality, Climate, and Economic Goals*. Cambridge, MA, US: MIT Press. ISBN: 978-0-262-01988-0 (↪ p. 28).

Nordhaus, William D. and Joseph Boyer (2000). "Warming the world. Economic models of global warming". In: MIT Press. ISBN: 0262140713 (↪ p. 22).

Oak Ridge National Laboratory (2016). *Transportation Energy Data Book*. ORNL-6990. URL: https://cta.ornl.gov/data/editions.shtml (↪ pp. 16, 29).

Parkes, Stephen D., Greg Marsden, Susan A. Shaheen, and Adam P. Cohen (2013). "Understanding the diffusion of public bikesharing systems: evidence from Europe and North America". In: *Journal of Transport Geography* 31, pp. 94–103. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2013.06.003 (↪ p. 17).

Pauw, W. P., Richard J. T. Klein, Kennedy Mbeva, Adis Dzebo, Davide Cassanmagnago, and Anna Rudloff (2018). "Beyond headline mitigation numbers: we need more transparent and comparable NDCs to achieve the Paris Agreement on climate change". In: *Climatic Change* 147.1, pp. 23–29. ISSN: 1573-1480. DOI: 10.1007/s10584-017-2122-x (↪ p. 18).

Pendakur, Krishna (2009). "EASI made Easier". In: *Quantifying Consumer Preferences*. Ed. by Daniel Jonathan Slottje. 1st ed. Contributions to Economic Analysis 288. Bingley, UK: Emerald. Chap. 7, pp. 179–206. ISBN: 9781848553125. DOI: 10.1108/S0573-8555(2009)0000288010 (↪ pp. 95, 96, 117, 128, 194).

Qing, Shen (2013). *Xinhua Insight: Tianjin's abrupt car-buying restriction raises doubts*. Xinhuanet. URL: http://news.xinhuanet.com/english/indepth/2013-12/18/c_132978558.htm (visited on 12/18/2015) (↪ p. 30).

Rausch, Sebastian, Gilbert E. Metcalf, and John M. Reilly (2011). "Distributional Impacts of carbon pricing: A general equilibrium approach with micro-data for households". In: *Energy Economics* 33.Supplement 1, S20–S33. DOI: 10.1016/j.eneco.2011.07.023 (↪ p. 22).

Rutherford, Thomas F. (1999). "Applied General Equilibrium Modeling with MPSGE as a GAMS Subsystem: An Overview of the Modeling Framework and Syntax". In: *Computational Economics* 14.1/2, pp. 1–46. DOI: 10.1023/A:1008655831209 (↪ p. 196).

Schäfer, Andreas, John B. Heywood, Henry D. Jacoby, and Ian A. Waitz (2009). *Transportation in a Climate-Constrained World*. Cambridge, MA, US: MIT Press. 356 pp. ISBN: 9780262512343 (↪ p. 16).

Sperling, Daniel and Deborah Gordon (2008). *Two Billion Cars: Drivings Towards Sustainability*. New York City, NY, US: Oxford University Press. ISBN:

9780195376647 (↪ p. 28).

Sussman, Joseph, Sgouris Sgouridis, and John Ward (2005). "New Approach to Transportation Planning for the 21st Century: Regional Strategic Transportation Planning as a Complex Large-Scale Integrated Open System". In: *Transportation Research Record. Journal of the Transportation Research Board* 1931, pp. 89–98. ISSN: 0361-1981. DOI: 10.3141/1931-12 (↪ p. 17).

Vandewalle, P., J. Kovacevic, and M. Vetterli (2009). "Reproducible research in signal processing". In: *IEEE Signal Processing Magazine* 26.3, pp. 37–47. ISSN: 1053-5888. DOI: 10.1109/MSP.2009.932122 (↪ p. 229).

Wang, Jiaoe, David Bonilla, and David Banister (2016). "Air deregulation in China and its impact on airline competition 1994–2012". In: *Journal of Transport Geography* 50. Air transport liberalisation and airline network dynamics: Investigating the complex relationships, pp. 12–23. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2015.03.007 (↪ pp. 16, 28).

Wang, Michael, Hong Huo, L. Johnson, and Dongquan He (2006). *Projection of Chinese Motor Vehicle Growth, Oil Demand and CO_2 Emissions Through 2050*. Tech. rep. ANL/ESD/06-6. Argonne, IL, US: Argonne National Laboratory, U.S. Department of Energy. URL: https://greet.es.anl.gov/publication-rwdz78ca (↪ pp. 28, 29).

Wang, Shenhao and Jinhua Zhao (2018). "Trajectories of Urban Development & Motorization: Clustering 287 Chinese Cities". Manuscript (in preparation) (↪ pp. 22, 164, 185).

Waraich, Rashid A., David Charypar, Michael Balmer, and Kay W. Axhausen (2015). "Performance Improvements for Large-Scale Traffic Simulation in MATSim". In: *Computational Approaches for Urban Environments*. Ed. by Marco Helbich, Jamal Jokar Arsanjani, and Michael Leitner. Vol. 13. Geotechnologies and the Environment. Cham, CH: Springer International Publishing, pp. 211–233. ISBN: 978-3-319-11469-9. DOI: 10.1007/978-3-319-11469-9_9 (↪ pp. 22, 23).

Wu, Tian, Mengbo Zhang, and Xunmin Ou (2014). "Analysis of Future Vehicle Energy Demand in China Based on a Gompertz Function Method and Computable General Equilibrium Model". In: *Energies* 7.11, pp. 7454–7482. ISSN: 1996-1073. DOI: 10.3390/en7117454 (↪ p. 24).

Xinhua News Agency (2014a). *China's Hangzhou city to restrict car ownership*. China Daily. URL: http://www.chinadaily.com.cn/china/2014-03/25/content_17378513.htm (visited on 12/18/2015) (↪ p. 30).

— (2014b). *Shenzhen restricts car purchases to ease congestion*. State Council Information Office. URL: http://www.china.org.cn/business/2014-12/29/content_34439091.htm (visited on 12/18/2015) (↪ p. 30).

Xue, Lulu (2015). *Four Lessons from Beijing and Shanghai Show How China's Cities Can Curb Car Congestion*. WRI Ross Center for Sustainable Cities. URL: http://www.wri.org/blog/2015/04/4-lessons-beijing-and-shanghai-show-how-china%E2%80%99s-cities-can-curb-car-congestion (visited on 12/18/2015) (↪ p. 30).

Yang, Jun, Ying Liu, Ping Qin, and Antung A. Liu (2014). *A Review of Beijing's Vehicle Lottery. Short-Term Effects on Vehicle Growth, Congestion, and Fuel Con-*

sumption. EfD Discussion Paper 14-01. URL: http://www.rff.org/research/publications/review-beijing-s-vehicle-lottery-short-term-effects-vehicle-growth-congestion (↪ p. 30).

Yang, Zifei and Anup Bandivadekar (2017). Tech. rep. Washington, DC, US: International Council on Clean Transportation. URL: https://www.theicct.org/publications/2017-global-update-LDV-GHG-FE-standards (↪ p. 28).

Yeh, Sonia, Gouri Shankar Mishra, Lew Fulton, Page Kyle, David L. McCollum, Joshua Miller, Pierpaolo Cazzola, and Jacob Teter (2016). "Detailed assessment of global transport-energy models' structures and projections". In: *Transportation Research Part D. Transport and Environment*. ISSN: 1361-9209. DOI: 10.1016/j.trd.2016.11.001 (↪ pp. 26, 27, 200).

Zeng, Heshuang (2012). *China Transportation Briefing: Stemming The Tide of Private Autos in Guangzhou*. WRI Ross Center for Sustainable Cities. URL: http://thecityfix.com/blog/china-transportation-briefing-stemming-the-tide-of-private-autos-in-guangzhou/ (visited on 12/18/2015) (↪ p. 30).

Zhang, Da, Sebastian Rausch, Valerie J. Karplus, and Xiliang Zhang (2013). "Quantifying regional economic impacts of CO_2 intensity targets in China". In: *Energy Economics* 40, pp. 687–701. DOI: 10.1016/j.eneco.2013.08.018 (↪ pp. 22, 47, 75).

Zhao, Jinhua and Zhongyue Wang (2014). "The Formation of a Transport Policy Market in China: From Policy Transfer to Policy Mobility". In: 93rd Annual Meeting of the TRB. (Washington, DC, US, Jan. 12, 2014). 14-4074. Transportation Research Board of the U.S. National Academies (↪ pp. 29, 195).

Zhao, Zhan, Jinhua Zhao, and Qing Shen (2013). "Has Shanghai's Transportation Demand Passed Its Peak Growth Period?" In: *Transportation Research Record. Journal of the Transportation Research Board* 2394: *Developing Countries 2013*, pp. 85–92. ISSN: 0361-1981. DOI: 10.3141/2394-11 (↪ p. 28).

Appendix A

Extended tables, listings & supplemental figures

A.1 Chapter 2

Table A.1: 2010, 2015, and 2030 energy-basis emissions factors from refined oil (`OIL`) combustion, by road transport subsector, province and species.

	Freight road (`FR`)			Household vehicle transport (`HVT`)		
Province	2010	2015	2030	2010	2015	2030
			BC [g/MJ]			
AH	11.5	3.34	1.78×10^{-2}	25.9	7.43	5.83×10^{-3}
BJ	3.21	0.959	1.78×10^{-2}	7.13	2.05	5.80×10^{-3}
CQ	11.7	3.41	1.78×10^{-2}	23.2	6.65	5.83×10^{-3}
FJ	8.28	2.42	1.78×10^{-2}	15.4	4.43	5.83×10^{-3}
GD	6.75	1.98	1.78×10^{-2}	14.7	4.22	5.83×10^{-3}
GS	12.9	3.75	1.78×10^{-2}	103	29.5	5.89×10^{-3}
GX	6.70	1.96	1.78×10^{-2}	17.5	5.02	5.83×10^{-3}
GZ	12.6	3.66	1.78×10^{-2}	27.3	7.85	5.84×10^{-3}
HA	8.41	2.45	1.78×10^{-2}	211	60.7	5.97×10^{-3}
HB	7.56	2.21	1.78×10^{-2}	82.2	23.6	5.87×10^{-3}
HE	13.6	3.94	1.78×10^{-2}	99.2	28.5	5.89×10^{-3}
HI	11.5	3.34	1.78×10^{-2}	13.3	3.82	5.83×10^{-3}
HL	12.4	3.58	1.78×10^{-2}	8.52	2.45	5.82×10^{-3}
HN	7.94	2.32	1.78×10^{-2}	28.3	8.14	5.84×10^{-3}
JL	9.78	2.84	1.78×10^{-2}	19.1	5.50	5.83×10^{-3}
JS	10.8	3.15	1.78×10^{-2}	45.8	13.2	5.85×10^{-3}
JX	7.34	2.14	1.78×10^{-2}	22.4	6.44	5.83×10^{-3}

Table A.1: (continued)

Province	Freight road (FR) 2010	2015	2030	Household vehicle transport (HVT) 2010	2015	2030
LN	9.37	2.73	1.78×10^{-2}	24.1	6.93	5.83×10^{-3}
NM	7.49	2.19	1.78×10^{-2}	40.1	11.5	5.84×10^{-3}
NX	8.06	2.35	1.78×10^{-2}	91.7	26.3	5.88×10^{-3}
QH	14.9	4.32	1.78×10^{-2}	78.5	22.5	5.87×10^{-3}
SC	8.41	2.45	1.78×10^{-2}	48.7	14.0	5.85×10^{-3}
SD	3.76	1.12	1.78×10^{-2}	34.3	9.84	5.84×10^{-3}
SH	14.7	4.25	1.78×10^{-2}	5.70	1.64	5.82×10^{-3}
SN	7.07	2.07	1.78×10^{-2}	13.6	3.91	5.83×10^{-3}
SX	13.6	3.93	1.78×10^{-2}	92.8	26.6	5.88×10^{-3}
TJ	7.69	2.24	1.78×10^{-2}	15.2	4.36	5.83×10^{-3}
XJ	6.50	1.90	1.78×10^{-2}	26.3	7.54	5.83×10^{-3}
YN	11.0	3.18	1.78×10^{-2}	24.6	7.06	5.83×10^{-3}
ZJ	8.90	2.59	1.78×10^{-2}	13.8	3.95	5.83×10^{-3}

Table A.1: (continued)

Province	Freight road (FR) 2010	2015	2030	Household vehicle transport (HVT) 2010	2015	2030
CO [g/MJ]						
AH	1130	325	2.12×10^{-1}	7120	2040	1.68×10^{-1}
BJ	522	150	2.11×10^{-1}	6550	1880	1.67×10^{-1}
CQ	1190	341	2.12×10^{-1}	7940	2280	1.68×10^{-1}
FJ	1000	287	2.12×10^{-1}	4350	1250	1.66×10^{-1}
GD	825	237	2.12×10^{-1}	3900	1120	1.65×10^{-1}
GS	1260	363	2.12×10^{-1}	23 600	6760	1.79×10^{-1}
GX	671	193	2.11×10^{-1}	3780	1080	1.65×10^{-1}
GZ	1330	381	2.12×10^{-1}	9670	2770	1.70×10^{-1}
HA	828	238	2.12×10^{-1}	43 400	12 500	1.94×10^{-1}
HB	765	220	2.12×10^{-1}	15 400	4420	1.74×10^{-1}
HE	1380	397	2.12×10^{-1}	28 100	8050	1.83×10^{-1}
HI	1220	351	2.12×10^{-1}	3220	923	1.65×10^{-1}
HL	1170	336	2.12×10^{-1}	3000	860	1.65×10^{-1}
HN	780	224	2.12×10^{-1}	7590	2180	1.68×10^{-1}
JL	974	280	2.12×10^{-1}	5250	1510	1.66×10^{-1}
JS	1050	303	2.12×10^{-1}	13 100	3760	1.72×10^{-1}
JX	710	204	2.12×10^{-1}	5450	1560	1.66×10^{-1}
LN	995	286	2.12×10^{-1}	7750	2220	1.68×10^{-1}
NM	671	193	2.11×10^{-1}	13 000	3730	1.72×10^{-1}
NX	823	236	2.12×10^{-1}	19 500	5600	1.77×10^{-1}
QH	1430	410	2.12×10^{-1}	17 100	4920	1.75×10^{-1}
SC	850	244	2.12×10^{-1}	23 100	6630	1.79×10^{-1}
SD	421	121	2.11×10^{-1}	10 200	2920	1.70×10^{-1}
SH	2330	669	2.13×10^{-1}	2890	828	1.65×10^{-1}
SN	676	194	2.11×10^{-1}	3960	1140	1.65×10^{-1}
SX	1320	380	2.12×10^{-1}	36 800	10 600	1.89×10^{-1}
TJ	909	261	2.12×10^{-1}	7200	2070	1.68×10^{-1}
XJ	655	188	2.11×10^{-1}	7660	2200	1.68×10^{-1}
YN	1140	327	2.12×10^{-1}	8410	2410	1.69×10^{-1}
ZJ	1110	318	2.12×10^{-1}	6070	1740	1.67×10^{-1}

Table A.1: (continued)

Province	Freight road (FR) 2010	Freight road (FR) 2015	Freight road (FR) 2030	Household vehicle transport (HVT) 2010	Household vehicle transport (HVT) 2015	Household vehicle transport (HVT) 2030
NO_X [g/MJ]						
AH	329	94.9	8.21×10^{-2}	502	144	7.77×10^{-3}
BJ	103	30.0	7.97×10^{-2}	339	97.4	7.60×10^{-3}
CQ	338	97.4	8.21×10^{-2}	533	153	7.79×10^{-3}
FJ	244	70.3	8.20×10^{-2}	238	68.2	7.58×10^{-3}
GD	204	58.9	8.20×10^{-2}	241	69.1	7.58×10^{-3}
GS	309	89.1	8.21×10^{-2}	1620	466	8.57×10^{-3}
GX	205	59.2	8.20×10^{-2}	201	57.6	7.56×10^{-3}
GZ	364	105	8.21×10^{-2}	673	193	7.89×10^{-3}
HA	234	67.7	8.20×10^{-2}	2760	792	9.38×10^{-3}
HB	214	61.9	8.20×10^{-2}	314	90.1	7.64×10^{-3}
HE	346	99.6	8.21×10^{-2}	1420	407	8.42×10^{-3}
HI	367	106	8.21×10^{-2}	209	59.9	7.56×10^{-3}
HL	280	80.8	8.21×10^{-2}	198	56.9	7.55×10^{-3}
HN	231	66.5	8.20×10^{-2}	482	138	7.76×10^{-3}
JL	230	66.4	8.20×10^{-2}	303	86.8	7.63×10^{-3}
JS	309	89.0	8.21×10^{-2}	627	180	7.86×10^{-3}
JX	216	62.5	8.20×10^{-2}	284	81.4	7.62×10^{-3}
LN	234	67.5	8.20×10^{-2}	537	154	7.80×10^{-3}
NM	175	50.6	8.20×10^{-2}	764	219	7.96×10^{-3}
NX	201	58.0	8.20×10^{-2}	1260	362	8.31×10^{-3}
QH	321	92.4	8.21×10^{-2}	1100	316	8.20×10^{-3}
SC	212	61.3	8.20×10^{-2}	1180	337	8.25×10^{-3}
SD	101	29.4	8.19×10^{-2}	474	136	7.75×10^{-3}
SH	733	211	8.24×10^{-2}	233	66.9	7.58×10^{-3}
SN	186	53.8	8.20×10^{-2}	237	68.1	7.58×10^{-3}
SX	347	99.9	8.21×10^{-2}	1860	534	8.74×10^{-3}
TJ	200	57.8	8.20×10^{-2}	414	119	7.71×10^{-3}
XJ	155	45.0	8.20×10^{-2}	489	140	7.76×10^{-3}
YN	321	92.5	8.21×10^{-2}	508	146	7.77×10^{-3}
ZJ	253	72.9	8.21×10^{-2}	390	112	7.69×10^{-3}

Table A.1: (continued)

Province	Freight road (FR) 2010	Freight road (FR) 2015	Freight road (FR) 2030	Household vehicle transport (HVT) 2010	Household vehicle transport (HVT) 2015	Household vehicle transport (HVT) 2030
OC [g/MJ]						
AH	4.40	1.30	1.78×10^{-2}	18.0	5.18	5.83×10^{-3}
BJ	1.31	0.414	1.78×10^{-2}	3.68	1.06	5.80×10^{-3}
CQ	4.52	1.33	1.78×10^{-2}	14.6	4.20	5.83×10^{-3}
FJ	3.45	1.03	1.78×10^{-2}	12.7	3.65	5.82×10^{-3}
GD	2.86	0.861	1.78×10^{-2}	12.1	3.47	5.82×10^{-3}
GS	4.81	1.42	1.78×10^{-2}	43.2	12.4	5.85×10^{-3}
GX	2.62	0.790	1.78×10^{-2}	12.7	3.64	5.82×10^{-3}
GZ	4.91	1.45	1.78×10^{-2}	14.7	4.22	5.83×10^{-3}
HA	3.21	0.960	1.78×10^{-2}	103	29.6	5.89×10^{-3}
HB	2.92	0.877	1.78×10^{-2}	83.4	23.9	5.88×10^{-3}
HE	5.16	1.52	1.78×10^{-2}	56.2	16.1	5.86×10^{-3}
HI	4.66	1.38	1.78×10^{-2}	9.66	2.78	5.82×10^{-3}
HL	4.54	1.34	1.78×10^{-2}	3.93	1.13	5.82×10^{-3}
HN	3.05	0.914	1.78×10^{-2}	18.8	5.41	5.83×10^{-3}
JL	3.66	1.09	1.78×10^{-2}	9.50	2.73	5.82×10^{-3}
JS	4.13	1.22	1.78×10^{-2}	37.6	10.8	5.84×10^{-3}
JX	2.82	0.847	1.78×10^{-2}	18.3	5.25	5.83×10^{-3}
LN	3.60	1.07	1.78×10^{-2}	12.6	3.63	5.82×10^{-3}
NM	2.73	0.823	1.78×10^{-2}	15.4	4.42	5.83×10^{-3}
NX	3.06	0.917	1.78×10^{-2}	39.5	11.4	5.84×10^{-3}
QH	5.45	1.60	1.78×10^{-2}	30.7	8.82	5.84×10^{-3}
SC	3.18	0.952	1.78×10^{-2}	32.6	9.35	5.84×10^{-3}
SD	1.51	0.473	1.78×10^{-2}	27.7	7.95	5.84×10^{-3}
SH	5.81	1.71	1.78×10^{-2}	3.51	1.02	5.82×10^{-3}
SN	2.66	0.801	1.78×10^{-2}	7.94	2.29	5.82×10^{-3}
SX	5.08	1.50	1.78×10^{-2}	57.8	16.6	5.86×10^{-3}
TJ	3.10	0.927	1.78×10^{-2}	9.36	2.69	5.82×10^{-3}
XJ	2.46	0.743	1.78×10^{-2}	12.8	3.67	5.82×10^{-3}
YN	4.25	1.26	1.78×10^{-2}	13.4	3.84	5.83×10^{-3}
ZJ	3.73	1.11	1.78×10^{-2}	7.91	2.28	5.82×10^{-3}

Table A.1: (continued)

Province	Freight road (FR) 2010	Freight road (FR) 2015	Freight road (FR) 2030	Household vehicle transport (HVT) 2010	Household vehicle transport (HVT) 2015	Household vehicle transport (HVT) 2030
SO_2 [g/MJ]						
AH	18.5	5.30	7.82×10^{-4}	18.5	5.33	7.43×10^{-3}
BJ	12.3	3.54	7.78×10^{-4}	12.4	3.56	7.37×10^{-3}
CQ	33.5	9.62	7.93×10^{-4}	33.6	9.65	7.44×10^{-3}
FJ	20.1	5.76	7.83×10^{-4}	20.1	5.79	7.43×10^{-3}
GD	16.8	4.82	7.81×10^{-4}	16.9	4.85	7.43×10^{-3}
GS	26.4	7.57	7.88×10^{-4}	26.4	7.60	7.43×10^{-3}
GX	25.8	7.40	7.87×10^{-4}	25.8	7.43	7.43×10^{-3}
GZ	31.3	8.99	7.91×10^{-4}	31.4	9.02	7.44×10^{-3}
HA	19.8	5.68	7.83×10^{-4}	19.8	5.71	7.43×10^{-3}
HB	26.3	7.55	7.88×10^{-4}	26.4	7.58	7.43×10^{-3}
HE	10.2	2.93	7.76×10^{-4}	10.2	2.96	7.42×10^{-3}
HI	14.2	4.08	7.79×10^{-4}	14.3	4.11	7.42×10^{-3}
HL	10.9	3.11	7.77×10^{-4}	10.9	3.14	7.42×10^{-3}
HN	27.0	7.74	7.88×10^{-4}	27.0	7.77	7.43×10^{-3}
JL	10.7	3.06	7.77×10^{-4}	10.7	3.09	7.42×10^{-3}
JS	20.2	5.80	7.83×10^{-4}	20.3	5.83	7.43×10^{-3}
JX	23.0	6.61	7.85×10^{-4}	23.1	6.63	7.43×10^{-3}
LN	16.4	4.70	7.81×10^{-4}	16.4	4.73	7.43×10^{-3}
NM	11.0	3.15	7.77×10^{-4}	11.0	3.18	7.42×10^{-3}
NX	25.7	7.37	7.87×10^{-4}	25.7	7.40	7.43×10^{-3}
QH	24.5	7.04	7.87×10^{-4}	24.6	7.07	7.43×10^{-3}
SC	17.4	4.98	7.81×10^{-4}	17.4	5.01	7.43×10^{-3}
SD	12.4	3.57	7.78×10^{-4}	12.5	3.60	7.42×10^{-3}
SH	18.5	5.31	7.82×10^{-4}	18.6	5.34	7.43×10^{-3}
SN	21.8	6.26	7.85×10^{-4}	21.8	6.28	7.43×10^{-3}
SX	11.5	3.31	7.77×10^{-4}	11.6	3.34	7.42×10^{-3}
TJ	12.5	3.59	7.78×10^{-4}	12.5	3.62	7.42×10^{-3}
XJ	21.6	6.21	7.84×10^{-4}	21.7	6.24	7.43×10^{-3}
YN	27.8	7.98	7.89×10^{-4}	27.9	8.01	7.43×10^{-3}
ZJ	17.5	5.03	7.82×10^{-4}	17.6	5.05	7.43×10^{-3}

A.2 Chapters 3 and 4

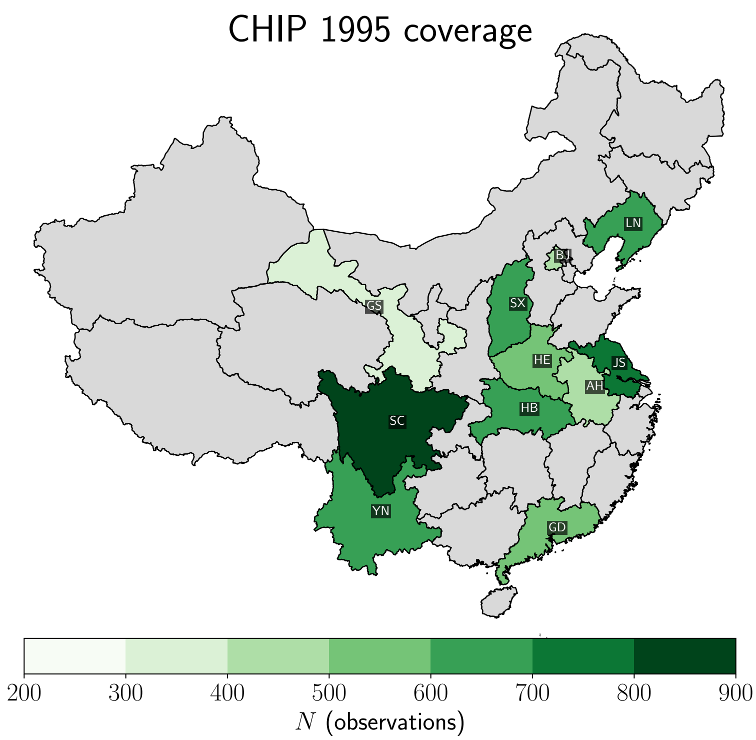


Figure A-1: Number of observations by province in the China Household Income Project (CHIP) 1995 cohort.

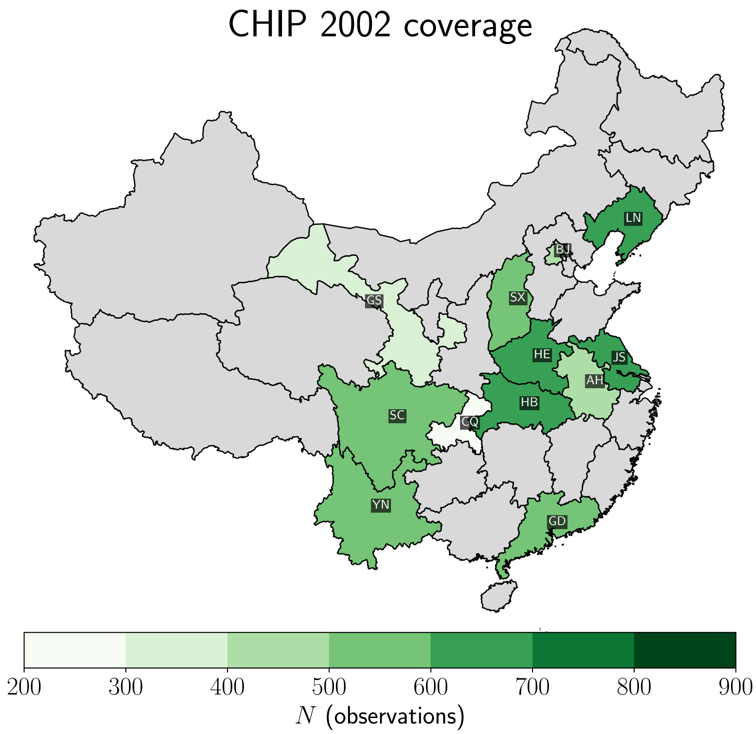


Figure A-2: Number of observations by province in the CHIP 2002 cohort.

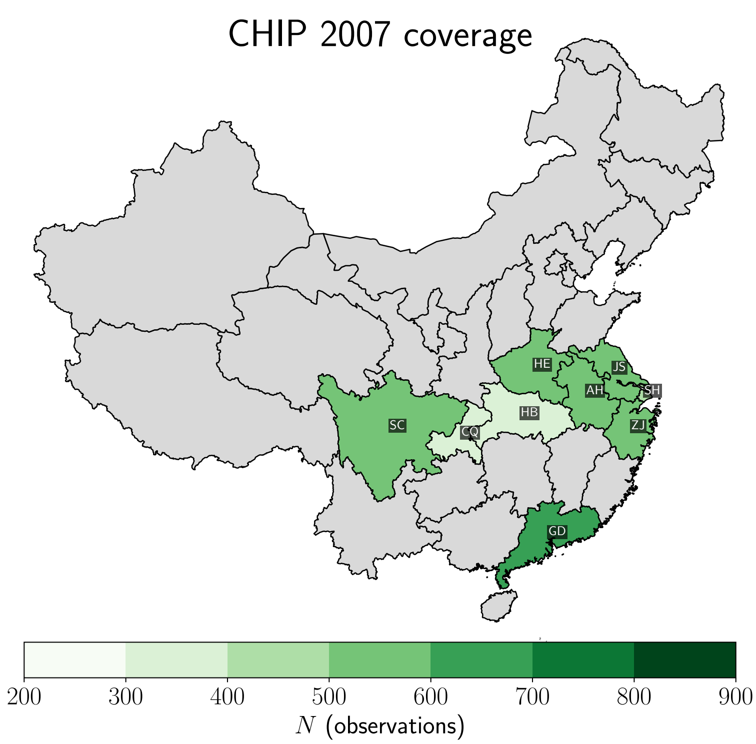


Figure A-3: Number of observations by province in the CHIP 2007 cohort.

Table A.2: Coverage of the CHIP survey data. Columns are: the 6-digit, county-level GB/T 2260 *code* describing the respondents' location; the number of household respondents in the region by CHIP wave; and an English name for the region, derived from the official database (see Appendix B.3). Where an English name is not available the name of the "parent" (higher-level) region is given. Some region codes have been reassigned over time.

Code	1995	2002	2007	Name
		北京市 Beijing		
110101		60		东城区 Dongcheng
110102		84		西城区 Xicheng
110103		53		
110104		60		
110105		78		朝阳区 Chaoyang
110106		35		丰台区 Fengtai
110107		34		石景山区 Shijingshan
110108		80		海淀区 Haidian
111100	492			
		山西省 Shanxi		
140100		200		太原市 Taiyuan
140200		93		大同市 Datong
140225		50		浑源县 Hunyuan
140400		99		长治市 Changzhi
140800		100		运城市 Yuncheng
141100	200			吕梁市 Luliang
141200	100			
141300	99			
141400	100			
142303		49		
142325		49		
143100	50			
143200	50			
143300	50			
		辽宁省 Liaoning		
210100		250		沈阳市 Shenyang
210200		249		大连市 Dalian
210281		50		瓦房店市 Wafangdian
210700		98		锦州市 Jinzhou
211100	299			盘锦市 Panjin
211200	200			铁岭市 Tieling
211224		50		昌图县 Changtu
211500	100			

Code	1995	2002	2007	Name
213100	50			
213200	50			
上海市 Shanghai				
310101			30	黄浦区 Huangpu
310103			14	
310104			57	徐汇区 Xuhui
310105			38	长宁区 Changningqu
310106			24	静安区 Jing'an
310107			57	普陀区 Putuoqu
310108			29	闸北区 Zhabei
310109			45	虹口区 Hongkouqu
310110			59	杨浦区 Yangpuqu
310112			28	闵行区 Minhang
310113			43	宝山区 Baoshan
310115			75	浦东新区 Pudongxin
江苏省 Jiangsu				
320100		149		南京市 Nanjing
320102			70	玄武区 Xuanwu
320103			80	
320104			50	秦淮区 Qinhuai
320105			50	建邺区 Jianyequ
320106			90	鼓楼区 Gulouqu
320107			59	
320200		99		无锡市 Wuxi
320202			42	崇安区 Chong'an
320203			67	南长区 Nanchang
320204			34	北塘区 Beitang
320211			57	滨湖区 Binhuqu
320282		48		宜兴市 Yixing
320300		97		徐州市 Xuzhou
320600		94		南通市 Nantong
320982		50		
321000		93		扬州市 Yangzhou
321100	200			镇江市 Zhenjiang
321200	100			泰州市 Taizhou
321283		49		泰兴市 Taixingshi
321300	99	50		宿迁市 Suqian
321500	100			
321900	100			
323100	50			
323200	50			

Code	1995	2002	2007	Name
323300	50			
323700	50			
		浙江省 Zhejiang		
330102			38	上城区 Shangcheng
330103			48	下城区 Xiacheng
330104			37	江干区 Jianggan
330105			41	拱墅区 Gongshu
330106			27	西湖区 Xihu
330108			4	滨江区 Binjiangqu
330109			50	萧山区 Xiaoshanqu
330110			41	余杭区 Yuhangqu
330182			100	建德市 Jiande
330203			54	海曙区 Haishu
330204			28	江东区 Jiangdong
330205			40	江北区 Jiangbei
330206			13	北仑区 Beilun
330211			39	镇海区 Zhenhai
330212			26	鄞州区 Yinzhouqu
		安徽省 Anhui		
340100		100		合肥市 Hefei
340102			91	瑶海区 Yaohaiqu
340103			125	庐阳区 Luyangqu
340104			84	蜀山区 Shushanqu
340111			50	包河区 Baohequ
340200		96		芜湖市 Wuhu
340300		99		蚌埠市 Bengbu
340302			72	龙子湖区 Longzihuqu
340303			75	蚌山区 Bangshanqu
340304			52	禹会区 Yuhuiqu
340400		100		淮南市 Huainan
341021		50		歙县 She
341100	100			滁州市 Chuzhou
341200	99			阜阳市 Fuyang
341400	100			
341500	100			六安市 Lu’an
341600		48		亳州市 Bozhou
343200	50			
343300	50			
		河南省 Henan		
410100		196		郑州市 Zhengzhou

Code	1995	2002	2007	Name
410102			70	中原区 Zhongyuan
410103			69	二七区 Erqi
410104			58	管城回族区 Guancheng Huizu
410105			90	金水区 Jinshui
410106			28	上街区 Shangjie
410108			29	惠济区 Huijiqu
410200		97		开封市 Kaifeng
410302			18	老城区 Laocheng
410303			56	西工区 Xigong
410304			21	河回族区 Chanhe Huizu
410305			67	涧西区 Jianxi
410306			14	吉利区 Jili
410307			21	
410400		99		平顶山市 Pingdingshan
410502			36	文峰区 Wenfeng
410503			25	北关区 Beiguan
410505			23	殷都区 Yinduqu
410506			16	龙安区 Longanqu
410526		48		滑县 Hua
410700		100		新乡市 Xinxiang
410782		50		辉县市 Huixian
411025		45		襄城县 Xiangchengxian
411100	100			漯河市 Luohe
411300	100			南阳市 Nanyang
411400	100			商丘市 Shangqiu
411500	99			信阳市 Xinyang
411525		45		固始县 Gushixian
413100	50			
413200	50			
413300	50			
413400	50			
				湖北省 Hubei
420100		244		武汉市 Wuhan
420102			48	江岸区 Jiang'an
420103			66	江汉区 Jianghan
420104			55	口区 Qiaokouqu
420105			37	汉阳区 Hanyang
420106			66	武昌区 Wuchang
420107			43	青山区 Qingshan
420111			43	洪山区 Hongshan
420500		91		宜昌市 Yichang
420600		98		襄阳市 Xiangyang

Code	1995	2002	2007	Name
421000		96		荆州市 Jingzhou
421083		49		洪湖市 Honghu
421100	283			黄冈市 Huanggang
421125		46		浠水县 Xishuixian
421200		49		咸宁市 Xianning
421300	100			随州市 Suizhou
421400	99			
422200	93			
423100	50			
423200	50			
423300	50			
		广东省 Guangdong		
440100		195		广州市 Guangzhou
440103			39	荔湾区 Liwan
440104			50	越秀区 Yuexiu
440105			50	海珠区 Haizhu
440106			50	天河区 Tianhe
440111			50	白云区 Baiyun
440112			50	黄埔区 Huangpu
440200		50		韶关市 Shaoguan
440303			40	罗湖区 Luohu
440304			50	福田区 Futian
440305			28	南山区 Nanshan
440306			29	宝安区 Bao’an
440307			29	龙岗区 Longgang
440308			20	盐田区 Yantianqu
440600		50		佛山市 Foshan
440681		49		
440800		50		湛江市 Zhanjiang
441100	193			
441200	48	50		肇庆市 Zhaoqing
441300	50	50		惠州市 Huizhou
441500	50			汕尾市 Shanwei
441700	50			阳江市 Yangjiang
441800	50			清远市 Qingyuan
441900			200	东莞市 Dongguan
443100	50			
443300	49			
445281		50		普宁市 Puningshi
		重庆市 Chongqing		
500100		196		市辖区 Shixiaqu

Code	1995	2002	2007	Name
500101		83		万州区 Wanzhouqu
500103			65	渝中区 Yuzhongqu
500104			30	大渡口区 Dadukouqu
500105			62	江北区 Jiangbeiqu
500106			40	沙坪坝区 Shapingbaqu
500107			37	九龙坡区 Jiulongpoqu
500108			20	南岸区 Nananqu
500109			33	北碚区 Beibeiqu
500112			58	渝北区 Yubeiqu
500113			38	巴南区 Bananqu
四川省 Sichuan				
510100		196		成都市 Chengdu
510104			58	锦江区 Jinjiang
510105			23	青羊区 Qingyang
510106			63	金牛区 Jinniu
510107			64	武侯区 Wuhou
510108			68	成华区 Chenghua
510109			22	
510500		97		泸州市 Luzhou
510703			148	涪城区 Fucheng
510704			51	游仙区 Youxian
510800		99		广元市 Guangyuan
511000		49		内江市 Neijiang
511100	199			乐山市 Leshan
511102			99	市中区 Shizhong
511181		50		峨眉山市 Emeishan
511200	200			
511300	97	94		南充市 Nanchong
511400	100			眉山市 Meishan
511600	100			广安市 Guang'an
511700	99			达州市 Dazhou
513500	50			
云南省 Yunnan				
530100		100		昆明市 Kunming
530200		98		
530381		50		宣威市 Xuanweishi
530500		94		保山市 Baoshan
531100	100			
531200	100			
531300	100			
531400	100			

Code	1995	2002	2007	Name
531500	98			
532501		95		个旧市 Gejiu
532722		50		
532901		99		大理市 Dali
533100	50			德宏傣族景颇族自治州 Dehong Daizu Jingpozu
533200	50			
533221		50		
533300	49			怒江傈僳族自治州 Nujiang
甘肃省 Gansu				
620100		198		兰州市 Lanzhou
621100	200			定西市 Dingxi
621300	99			
621500	100			
622301		97		
622701		100		

A.2.1 Model specifications

```
# Model specifications

1:
  name: 'y5-{city}'
  power: 5
  hh: [gender, age, educ, single]
  city:
    # One-at-a-time city-level variables
    # Primary data
    only-gdp_cap:             [gdp_cap            ]
    only-wage_avg:            [wage_avg           ]
    # Derived
    only-density:             [density            ]
    only-gdp_density:         [gdp_density        ]
    only-hwy_density:         [hwy_density        ]
    only-p_hwy_cap:           [p_hwy_cap          ]
    only-stock_bus_cap:       [stock_bus_cap      ]
    only-stock_priv_cap:      [stock_priv_cap     ]
    only-stock_rent_cap:      [stock_rent_cap     ]
    # Rural data only
    # only-area_floor:        [area_floor         ]
    # Only available at the provincial level
    # only-p_rail_cap:        [p_rail_cap         ]
```

```
    # only-road_density:      [road_density       ]
    # only-stock_comm_pass_cap: [stock_comm_pass_cap]
    # only-stock_priv_pass_cap: [stock_priv_pass_cap]
    # Prices
    only-p_trn_fac:           [p_trn_fac          ]
    only-p_trn_fuel:          [p_trn_fuel         ]
    only-p_trn_ic:            [p_trn_ic           ]
    only-p_trn_maint:         [p_trn_maint        ]
    only-p_trn_pt:            [p_trn_pt           ]
    # Sensitivity check
    density+gdp_cap:          [density, gdp_cap   ]
    # All at once
    many-dem:
      - gdp_cap  # Primary
      - wage_avg
      - density  # Derived
      - gdp_density
      - hwy_density
      - p_hwy_cap
      - stock_bus_cap
      - stock_priv_cap
      - stock_rent_cap
      - p_trn_fac  # Prices
      - p_trn_fuel
      - p_trn_ic
      - p_trn_maint
      - p_trn_pt
  options: [province year fixed effects]


2:  # Various sensitivity & feature checks
  name: 'test{options}'
  power: 5
  hh: []
  city:
    - gdp_cap
    - wage_avg
    - density
    - hwy_density
    - p_hwy_cap
    - stock_bus_cap
    - stock_priv_cap
    - stock_rent_cap
    - p_trn_fuel
  options:
    '': []
    +nocensor:
```

```
      censor: False
    # a dummy in years where Shanghai had the vehicle license plate policy
    +vlp: [vlp dummy]
    # fixed effects
    +prov_fe: [province fixed effects]
    +year_fe: [year fixed effects]
    +provyear_fe: [province year fixed effects]
    # use logarithms of city variables → elasticities
    +logs: [log dem]

# Main models

3: &master # master group
  name: 'y{power}{hh}{city}'
  power:
    1: 1
    3: 3
    5: 5
    6: 6
  hh:
    '': []
    '+hh': [gender, age, educ, single]
  city:
    '': []
    '+city':
      - gdp_cap
      - wage_avg
      - density
      - hwy_density
      - p_hwy_cap
      - stock_bus_cap
      - stock_priv_cap
      - stock_rent_cap
      - p_trn_fuel

# variants of master group with fixed effects
4:
  <<: *master
  name: 'y{power}{hh}{city}+fe_p'
  options: [province fixed effects]

5:
  <<: *master
  name: 'y{power}{hh}{city}+fe_py'
  options: [province year fixed effects]

6:
```

```
  <<: *master
  name: ’y{power}{hh}{city}+fe_y’
  options: [year fixed effects]

# for infill of groups {36
7:
  <<: *master
  name: ’y{power}{hh}{city}{options}’
  power:
    2: 2
    4: 4
    7: 7
  options:
    ’’: []
    ’+fe_p’: [province fixed effects]
    ’+fe_py’: [province year fixed effects]
    ’+fe_y’: [year fixed effects]

# For paper 014
10:
  <<: *master
  name: ’y4{hh}{city}{options}’
  power: 4
  options:
    ’’: []
    ’+fe_p’: [province fixed effects]
    ’+fe_py’: [province year fixed effects]
    ’+fe_y’: [year fixed effects]


8: # the AIDS model. No demographic variables, so no options
  name: ’aids’
  type: aids


9: # simple models for comparing with AIDS
  name: ’{power}’
  power:
    linear: 1
    quadratic: 2
```

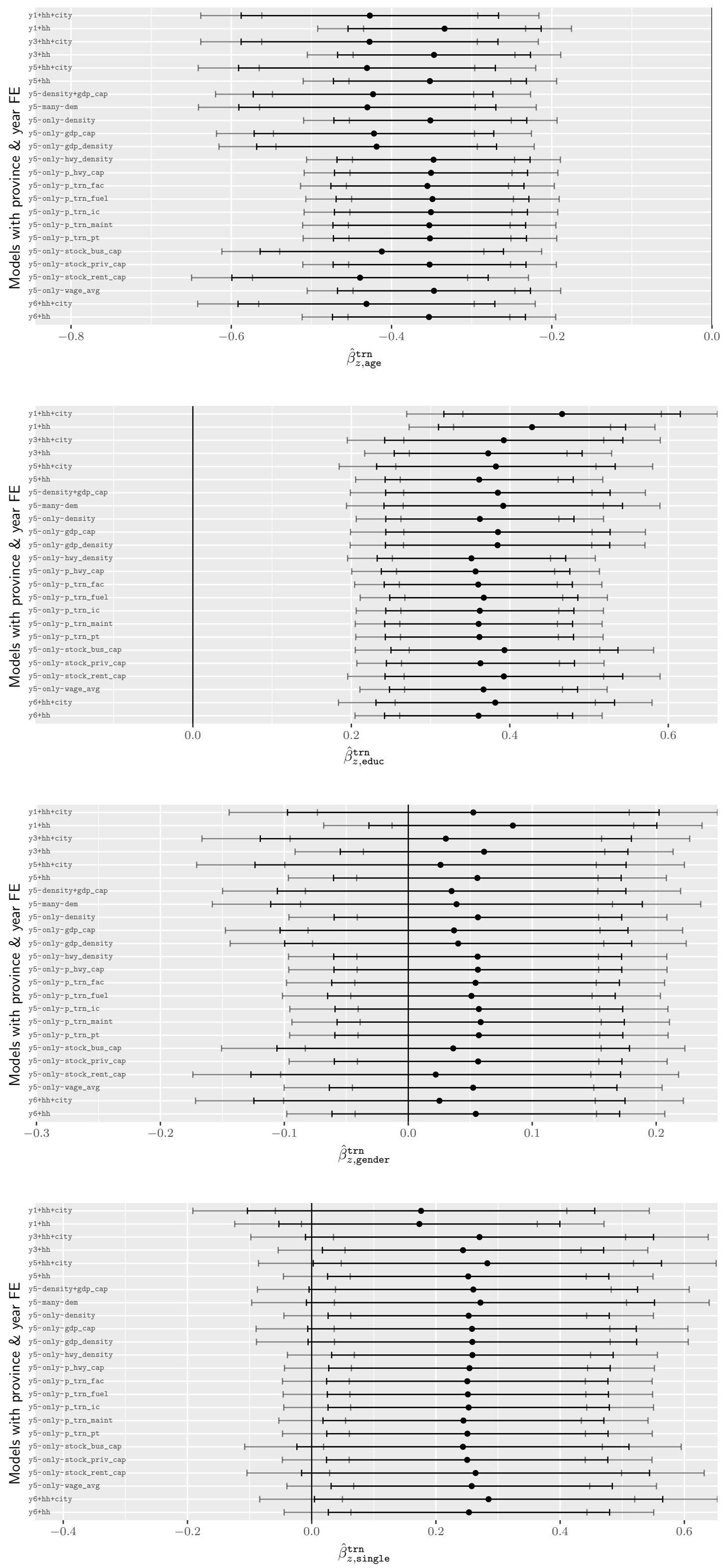


Figure A-4: Parameter estimates for household variables on $w^{\mathtt{trn}}$

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix B

Materials for reproduction

To enable reproduction (Vandewalle et al. 2009) and extension of the work in this thesis, software and materials for some of the foregoing is available, as described in this appendix. Software is provided in two forms at https://paul.kishimoto.name/publications:

1. Public version control repositories. These may undergo further development in response to peer review of papers proceeding from the thesis or as part of subsequent research.
2. Zenodo archives. These have Digital Object Indentifiers (DOIs), are citable, and refer to specific versions of the code from #1.

B.1 Core models, analysis, and presentation

In Chapter 2, the C-REM model is property of the Tsinghua-MIT China Energy and Climate Project (CECP)—a joint effort of the MIT Joint Program on the Science and Policy of Global Change (JPSPGC) and the Tsinghua University Institute of Energy, Environment, and Economy (3E). The model source is not public; but can be provided on request, under certain restrictions on publication and re-use. Interested parties should contact the JPSPGC:

Online https://globalchange.mit.edu/about-us/our-purpose/contact-us

Telephone +1 617-253-7492

Post Joint Program on the Science and Policy of Global Change

Massachusetts Institute of Technology

77 Massachusetts Ave, E19-411

Cambridge, MA 02139-4307, USA

Analysis and plotting code for Chapter 2, as well as model estimation, validation, and analysis code for Chapters 3 and 4 are provided at `https://paul.kishimoto.name/publications`. `https://github.com/khaeru/easi` contains an improved version of the exact affine Stone index (EASI) estimation code in R developed by Hoareau et al. (2012).

B.2 Data and software for data preparation

This software is available at `https://github.com/khaeru/data`. Specifically, the subpackages `chip`, `ceic` and `cn_nbs` are used, each described below. These Python packages retrieve raw data from raw or original data files, or the Internet; merge it; apply cleaning steps, transformations (of region or variable identifiers), and consistency checks; and cache the resulting data structures. Each provides methods with names like `load_chip()` that allow extraction of the resulting data in formats convenient for modeling and analysis.

B.2.1 China Household Income Project (CHIP)

The China Household Income Project (CHIP) data files are available from `http://www.ciidbnu.org/chip/index.asp?lang=EN` free of charge. The researchers require prospective users to complete a short application for access; in order to not circumvent this, I have not republished the raw data. The data consist of `.rar` archives containing Stata `.dta` files and accompanying documentation, organized by three dimensions, each with a string label. The software uses following names for these dimensions:

`sample` e.g. 'rural', 'urban' or 'migrant'.

`unit` of measurement/analysis: e.g. 'person' or 'household'.

`section` of a questionnaire or survey: e.g. 'abc' or 'income and assets'.

Some CHIP surveys contain multiple sections which apply to individual people, but the units of observation are distinct: for instance, all household members; children of the household head(s) who are not resident in the household; or parents of the household head(s) who may or may not be resident in the household. These are usually indexed by different variables and have different data associated with them, so the code stores them separately.

The code is tested for CHIP waves between 1988 and 2013 inclusive. Since the provided form of the data varies from wave to wave, we create metadata/control files that describes the layout of the data, units of observations, and/or columns to be transformed or used as indices. The file for the 2002 CHIP wave is reproduced below on page 234.

B.2.2 CEIC Data

China's official economic and transport statistics are collected by a system of institutions, then aggregated, collected, and finally published by several of these. At the central level are the National Bureau of Statistics of China (NBSC) and the Ministry of Transport. Each province has its own statistical bureau, transport ministry, and offices of the national bodies. City governments also have transport departments.

Data collected by these institutions is reported annually in the national-level (general) China Statistical Yearbook, and (sector-specific) China Transportation & Communications Yearbook, and at other frequencies through the NBSC website. The province-level institutions issue publications such as a "Jiangsu Statistical Yearbook"—analogous to the general, national-level yearbook—and sector-specific provincial yearbooks. These contain tables at resolutions below the province level, but their contents are not published electronically alongside the digital form of the national yearbooks.

This thesis uses the "China Premium Database" published by CEIC Data (CEIC), a commercial source which has transcribed a large number of such yearbooks. While this saves researchers the effort of collation, the data available from the CEIC 'CDM-Next' platform is not immediately suitable for analysis of the type described in Chapters 3 and 4. In particular, data series are tagged with string names such as the following:

— `No of Public Transit Vehicle: Bus and Trolley Bus: Henan: Hebi`

— `GDP: Shanxi: Changzhi: Zhangzi`

— `Highway: Freight Traffic: Commercial: Motor Vehicle`

These names combine measured concepts (e.g. 'GDP'), subcategories of measurement ('Commercial' within 'Freight Traffic'), and names of geographical units: in the examples, respectively the prefecture-level city 'Hebi'; the county-level city district 'Zhangzhi'; and, implicitly, all of China. The names names, in particular, contain errors, use idiosyncratic, non-official romanizations for names in China's minority languages and sometimes for Mandarin, and refer to regions that have been renamed over the duration of the series they label.

The provided software in the `data` repository systematizes the CEIC CDMNext exports more carefully. Among other transformations, it:

— groups indicators and sub-indicators,

— flexibly matches names to the official 6-digit GB/T 2260 codes for geographical divisions, using the library described in Appendix B.3,

— converts series in heterogeneous units to common units, and

— replaces series codes such as `'CTBBZV'` with intelligible names like `'stock_bus'`.

The file reproduced on page 236 encodes the corrections made at some of these stages. The software, control file, and CEIC series exports in comma-separated value (CSV) format can be used to reproduce data for the city-level measures constructed by the software in Appendix B.1.

B.2.3 China National Bureau of Statistics

The NBSC website provides price indices for a general basket, for eight top-level categories (the same used in Chapters 3 and 4), further subcategories, and for the individual goods whose prices are directly surveyed in order to construct the indices.

These data are provided online through an interactive website that allows export of small subsets. The provided software in the `data` repository retrieves entire series in JSON format through the web application programming interface (API) that backs the website. Raw data are cached, converted to Python data structures, and cached again in this format.

B.3 Utilities and presentation

Two new utility packages were developed for this thesis. I also gratefully acknowledge the efforts of the Python and Free Software community members who developed and supported the `pandas`, `xarray`, `rpy2`, `matplotlib`, `plotnine`, and `statsmodels` libraries.

GB/T 2260-2007: Codes for the divisions of the People's Republic of China

https://github.com/khaeru/gb2260

Officially "中华人民共和国行政区划代码," the GB/T 2260 standard defines six-digit numerical codes for the administrative divisions of China, at the county level and above. For instance, the Haidian district of Beijing has the code 110108. The most recent version of the official standard, designated "GB/T 2260-2007," was published in 2008. However, codes are routinely revised, and the National Bureau of Statistics (NBS) publishes an updated list online annually.

The gb2260 package produces and exposes an up-to-date list of the GB/T 2260 codes, with extra information including English names, Pinyin transcriptions, administrative levels, etc. It conforms to the semantics of the widely used pycountry package, an interface to the ISO 3166 family of standards.

Pandas tables in LaTeX

https://github.com/khaeru/pandas_latex

This package produces LaTeX tables from the popular pandas Python data structures. The analysis code in Appendix B.1 uses this library to generate most of the data tables in this thesis.

B.4 Listings

B.4.1 Metadata for CHIP 2002 wave

```
files:
  path: 2002
  name: 21741-00(?P<sample>[01][0-9])-Data.dta

map:
  _dim: sample
  '01':  # Must be quoted or else are converted to 1 != '01'
    sample: urban
    unit: person
    section: Income, consumption and employment
  '02':
    sample: urban
    unit: household
    section: Income, consumption and employment
  '03':
```

```
    sample: urban
    unit: person
    section: Annual income
  '04':
    sample: urban
    unit: household
    section: Assets, expenditure, income and conditions
  '05':
    sample: rural
    unit: village
    section: Administrative
  '06':
    sample: rural
    unit: r_person
    section: Income, consumption and employment
  '07':
    sample: rural
    unit: r_household
    section: Income, consumption, employment, social network, quality of life, village affa
  '08':
    sample: rural
    unit: r_person
    section: School-age children
  '09':
    sample: migrant
    unit: m_person
    section: All
  '10':
    sample: migrant
    unit: m_household
    section: All

column:
  PCODE:
    name: Household member code
    type: int
  CODE_P:
    name: Household member code
    type: int

unit:
  household:
    index: PCODE
  person:
    index: [CITY, PCODE, CODE_P]
    unique: false
  m_household:
```

```
    index: CODE
  m_person:
    index: [CODE, P102]
  r_household:
    index: [COUN, VILL, HOUS]
  r_person:
    index: [COUN, VILL, HOUS, P1_2]
  village:
    index: [COUN, VILL]
```

B.4.2 Metadata for CEIC Data

```
# Translations for CEIC data base city names which do not match official
# names. Mostly these different romanizations for names in minority languages.
rename regions:
  Aba: Aba Zangzu Qiangzu  # Sichuan (to disambiguate with its subdivision)
  Aksu: Aksu diqu  # Xinjiang (to disambiguate with its subdivision)
  Altay: Altay diqu  # Xinjiang (to disambiguate with its subdivision)
  Aletai: Altay  # Altay, Xinjiang
  Atushi: Artux shi  # Kizilsu Kirgiz, Xinjiang
  Baxiu: Basu  # Qamdu, Tibet
  Biyang:  Miyang  # Zhumadian, Henan
  Botou: Potou  # Cangzhou, Hebei
  Buerjin: Burqin  # Altay, Xinjiang
  Chancheng:  Shancheng  # Foshan, Guangdong
  Chengduo: Chenduo  # Yushu, Qinghai
  Dacheng: Daicheng  # Langfang, Hebei
  Danling: Danleng  # Meishan, Sichuan
  Deqin: Deqen  # Diqing, Yunnan
  Donge: Donga  # Liaocheng, Shandong
  Dujun: Duyun  # South Guizhou, Guizhou
  Erdos: Ordos  # Inner Mongolia
  Fanshi: Fanzhi  # Xinzhou, SX
  Fuijian: Fujian
  Fuxin Mongolian:  Fuxin Mongolzu  # Fuxin, Liaoning
  Hami: Hami diqu  # Xinjiang (to disambiguate with its subdivision)
  Hetian City: Hetian shi  # Hotan, Xinjiang
  Honghe: Honghe Ha'nizu Yizu  # Yunnan (to disambiguate with its subdivision)
  Jizhou: Ji # Tianjin
  Junlian: Yunlian  # Yibin, Sichuan
  Kashi: Kaxgar  # Xinjiang
  Kazuo:  Kalaqin zuoyi  # Chaoyang, Liaoning
  Laoting: Leting  # Tangshan, Hebei
  Lhokha: Loka  # Tibet
  Lindian:  Lindain  # Daqing, Heilongjiang
  Lingtao: Lintao  # Dingxi, Gansu
```

```
  Linxia City: Linxia shi  # Linxia, Gansu
  Liuzhi:  Luzhi  # Liupanshui, Guizhou
  Lueyang: Lveyang # Hanzhong, Shaanxi (error in gb2260)
  Lvchun: Luchun  # Honghe, Yunnan
  Lvliang: Luliang  # Shanxi
  MolidawaDaur:  Molidawadawoer  # Hulunbeier, Inner Mongolia
  Muling:  Muleng  # Mudanjiang, Heilongjiang
  Nanmulin: Namling  # Xigaze, Tibet
  Narqu: Nagqu  # Tibet
  Ngri: Ngari  # Tibet
  Pizhou:  Peizhou  # Xuzhou, Jiangsu
  Sahuangjiang: Shuangjiang  # Lincang, Yunnan
  Sui: Suixian  # Shangqiu, Henan (to disambiguate w/ Suiyangqu, Shangqiu)
  Suiling:  Suileng  # Suihua, Heilongjiang
  Tacheng: Tacheng diqu  # Xinjiang (to disambiguate with its subdivision)
  Tanchang: Dangchang  # Longnan, Gansu
  Wuchuang: Wuchuan  # Zunyi, Guizhou
  Wulumuqi: Urumqi  # 650121, in Urumqi, Xinjiang
  Xian: Xi'an  # Shaanxi (to disambiguate w/ Xianyang, Shaanxi)
  Xilinggol: Xilingol  # Inner Mongolia
  Xun: Jun  # Hebi, Henan
  Zhalaite:  Zalaite  # Xingan, Inner Mongolia
  Zhashui: Zuoshui  # Shangluo, Shaanxi
  Zhong: Zhongxian  # Chongqing (to disambiguate with Zhongqingshi)
  Zhongmu: Zhongmou  # Zhengzhou, Henan
  Zhongweishixiaqu: Zhongwei city area  # Zhongwei, Ningxia (error in gb2260)


# The following appear in some series names, but not in the 2015 gb2260 database
missing regions:
  "Tianjin:Baodi": 0
  "Shanxi:Yangquan:Meng": 0
  "Hebei:Shijiazhuang:Xinji": 139002  # was 130181, now not in Shijiazhuang
  "Hebei:Baoding:Li": 130635  # ambiguous with 130606 Lianchiqu
  "Inner Mongolia:Erdos:Etuoke": 150624  # ambiguous w/ 150623 Etuokeqianqi
  "Liaoning:Jinzhou:Linhai": 210781  # ambiguous with 331082
  "Anhui:Huainan:Shou": 0
  "Anhui:Xuancheng:Jing": 341823  # ambiguous with 341825 Jingdexian
  "Jiangxi:Shangrao:Yanshan": 0
  "Shandong:Heze:Shan": 0
  "Hubei:Shiyan:Yun": 420304  # Yunyangqu
  "Hubei:Qianjiang": 429005  # doesn't match at 2nd level
  "Guangdong:Meizhou:Mei": 441402  # ambiguous with 610326
  "Guangdong:Yunfu:Yunan": 0  # CEIC doesn't distinguish 445302 Yunanqu and
                              # 445322 Yunanxian
  "Hainan:Sanya:Baisha": 0
  "Sichuan:Liangshan:Huidong": 513426  # Liangshan matches 370832 (bug)
```

```
  "Sichuan:Liangshan:Yuexi": 513434
  "Guizhou:Southwest Guizhou:Jinsha": 520523  # under Bijie, not 522300
  "Yunnan:Dehong:Luxi": 532527  # under Honghe, not 533100 Dehong
  "Gansu:Longnan:Li": 621226  # can't rename bc 'Li' appears elsewhere
  "Gansu:Linxia:Linxia County": 622921
  "Xinjiang:Hotan:'Hetian County'": 653221
  "Xinjiang:Ili Kazak:Yining": 0  # CEIC doesn't distinguish 654002 Yiningshi
                                  # and 654021 Yiningxian


dimensions:
  airport:
    - [Airport, Freight Throughput]
    - [Airport, Passenger Throughput]
    - [Airport, No of Flight Handled]
  brand:
    - [New Registration of Passenger Car, by Brand]


preprocess:
  - predicate: "'Petroleum Product' in row['name']"
    transform: "row['name'] = row['name'][1:] + row['name'][:1]"


rename variables:  # Internal name: [CEIC name fragment]
  area: [Land Area of Administrative Zone]
  area_city: [Developed Area of City Construction]
  area_floor: [Floor Area of Residential Building per Capita]
  area_floor: [Floor Area of Residential Building per Capita, Rural]
  area_road: [Area of Paved Road, City]
  exp_cap: [Consumption Expenditure per Capita]
  exp_cap_rural2: [Living Exp per Capita, Rural Household]
  exp_cap_rural: [Consumption Expenditure per Capita, Rural]
  f_air: [Airport, Freight Throughput]
  f_all: [Transport, Freight Traffic]
  f_hwy: [Highway, Freight Traffic]
  f_hwy_comm: [Highway, Freight Traffic, Commercial]
  f_hwy_comm_mv: [Highway, Freight Traffic, Commercial, Motor Vehicle]
  f_hwy_comm_omv: [Highway, Freight Traffic, Commercial, Other Motor Vehicle]
  f_hwy_comm_trac: [Highway, Freight Traffic, Commercial, Tractor]
  f_rail: [Railway, Freight Traffic]
  gdp: [GDP]
  gdp_cap: [GDP, per Capita]
  goods_rural_bike: [Consumer Goods per 100 Rural Household, Bicycle]
  goods_rural_mc: [Consumer Goods per 100 Rural Household, Motor Cycle]
  goods_urban_auto: [Consumer Goods per 100 Urban Household, Automobile]
  hh: [No of Household]
```

hwy: [Highway, Length of Highway]
hwy_all: [Highway, Length of Highway, Expressway & Class I to IV]
hwy_c1: [Highway, Length of Highway, Class I]
hwy_c2: [Highway, Length of Highway, Class II]
hwy_c3: [Highway, Length of Highway, Class III]
hwy_c4: [Highway, Length of Highway, Class IV]
hwy_expwy: [Highway, Length of Highway, Expressway]
inc_cap: [Disposable Income per Capita]
oil_cons: [Petroleum Product, Consumption]
oil_fcons: [Petroleum Product, Final Consumption]
oil_fcons_res: [Petroleum Product, Final Consumption, Residential]
oil_fcons_res_rural:
  - Petroleum Product
  - Final Consumption
  - Residential
  - Rural
oil_fcons_res_urban:
  - Petroleum Product
  - Final Consumption
  - Residential
  - Urban
oil_fcons_tran:
  - Petroleum Product
  - Final Consumption
  - Transport, Storage, Postal & Telecommunication Service
p_air: [Airport, Passenger Throughput]
p_all: [Transport, Passenger Traffic]
p_hwy: [Highway, Passenger Traffic]
p_hwy_comm: [Highway, Passenger Traffic, Commercial]
p_hwy_comm_mv: [Highway, Passenger Traffic, Commercial, Motor Vehicle]
p_hwy_pt: [Highway, Passenger Traffic, Public Transport]
p_rail: [Railway, Passenger Traffic]
price_prop: [Property Price]
price_prop_res: [Property Price, Residential]
pkm_coastal: [Coastal, Passenger Turnover]
pkm_hwy: [Highway, Passenger Turnover]
pkm_hwy_comm: [Highway, Passenger Turnover, Commercial]
pkm_hwy_comm_mv: [Highway, Passenger Turnover, Commercial, Motor Vehicle]
pkm_ocean: [Ocean, Passenger Turnover]
pkm_river: [River, Passenger Turnover]
pkm_ww: [Waterway, Passenger Turnover]
pop: [Population]
pop_census: [Population, Census]
pop_non_ag: [Population, Non Agricultural]
stock_bus: [No of Public Transit Vehicle, Bus and Trolley Bus]
stock_comm: [No of Motor Vehicle, Commercial]
stock_comm_pass: [No of Motor Vehicle, Commercial, Passenger]

```
  stock_comm_truck: [No of Motor Vehicle, Commercial, Truck]
  stock_comm_truck_gen: [No of Motor Vehicle, Commercial, Truck, General]
  stock_comm_truck_spec:
    - No of Motor Vehicle
    - Commercial
    - Truck
    - Special Purpose
  stock_mc: [No of Motorcycle]
  stock_mc_priv: [No of Motorcycle, Private Owned]
  stock_mc_priv_2w: [No of Motorcycle, Private Owned, Two Wheelers]
  stock_other: [No of Motor Vehicle, Other Type]
  stock_pass: [No of Motor Vehicle, Passenger]
  stock_pass_large: [No of Motor Vehicle, Passenger, Large]
  stock_priv: [No of Motor Vehicle, Private Owned]
  stock_priv_pass: [No of Motor Vehicle, Private Owned, Passenger]
  stock_priv_pass_large: [No of Motor Vehicle, Private Owned, Passenger, Large]
  stock_priv_truck: [No of Motor Vehicle, Private Owned, Other Type]
  stock_priv_truck: [No of Motor Vehicle, Private Owned, Truck]
  stock_rent: [No of Rental Vehicle]
  stock_trac: [No of Tractor]
  stock_trac_priv: [No of Tractor, Private Owned]
  stock_truck: [No of Motor Vehicle, Truck]
  tkm_coast: [Coastal, Freight Turnover]
  tkm_hwy_comm: [Highway, Freight Turnover, Commercial]
  tkm_hwy_comm_mv: [Highway, Freight Turnover, Commercial, Motor Vehicle]
  tkm_hwy_comm_omv: [Highway, Freight Turnover, Commercial, Other Motor Vehicle]
  tkm_hwy_comm_trac: [Highway, Freight Turnover, Commercial, Tractor]
  tkm_ocean: [Ocean, Freight Turnover]
  tkm_ww: [Waterway, Freight Turnover]
  wage_avg: [Average Wage]
  wage_avg_duty: [Average Wage, On Duty]
  wage_tl: [Total Wage]
  wage_tl_duty: [Total Wage, On Duty]


units: |
  percent = [percent]
  person = [person]
  RMB = [currency]
  unit = [unit]
  jan2004 = 100

  cub_m_mn = 1e6 * m ** 3
  ha_th = 1e3 * hectare
  meter_th = 1e3 * metre
  person_km_mn = 1e6 * person * km
  person_mn = 1e6 * person
```

```
person_th = 1e3 * person
sq_km = km ** 2
sq_m = m ** 2
sq_m_mn = 1e6 * m ** 2
rmb = RMB
rmb_mn = 1e6 * RMB
rmb_bn = 1e9 * RMB
time_mn = person_mn
ton_km_mn = 1e6 * tonne * km
ton_mn = 1e6 * tonne
ton_th = 1e3 * tonne
unit_th = 1e3 * unit
```