Essays in Public Economics and Political Economy

by

Maxim L. Pinkovskiy

B.A., Economics-Mathematics, Columbia University (2008)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

Signature of Author ....
Department of Economics
15 May 2013

Certified by ....
Daron Acemoglu
Elizabeth and James Killian Professor of Economics
Thesis Supervisor

Certified by ....
Jerry Hausman
John and Jennie S. MacDonald Professor of Economics
Thesis Supervisor

Accepted by ....
Michael Greenstone
3M Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Essays in Public Economics and Political Economy
by
Maxim L. Pinkovskiy

Submitted to the Department of Economics
on 15 May 2013, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis studies topics in public economics in developed and developing countries, including health insurance regulation, public goods provision and inequality and welfare measurement. The first chapter analyzes the impacts of the managed care backlash in the United States on health care costs in the late 1990s and early 2000s. During the late 1990s, most U.S. states passed a variety of laws in this period that restricted the cost-cutting measures that managed care organizations (HMOs, PPOs and others) could use. I exploit panel variation in the passage of these regulations across states and over time to investigate the effects of the managed care backlash, as proxied by this legislation, on health care cost growth. I find that the backlash had a strong effect on health care costs, and can statistically explain much of the rise in health spending as a share of U.S. GDP between 1993 and 2005 (amounting to 1% - 1.5% of GDP). I also investigate the effects of the managed care backlash on intensity of care, hospital salaries and technology adoption. I conclude that managed care was largely successful in keeping health care costs on a sustainable path relative to the size of the economy. The second chapter attempts to quantify the impact of differences in political factors on economic growth and development, and specifically, assess to what extent variation in public goods provision may be responsible for cross-country differences in income and growth rates. Using a new methodology for the computation of standard errors in a regression discontinuity design with infill asymptotics, I document the existence of discontinuities in the levels and growth of the amount of satellite-recorded light per capita across national borders. Both the amount of lights per capita and its growth rate are shown to increase discontinuously upon crossing a border from a poorer (or lower-growing) into a richer (or higher-growing) country. I argue that these discontinuities form lower bounds for discontinuities in economic activity across borders, which suggest the importance of national-level variables such as institutions and culture relative to local-level variables such as geography for the determination of income and growth. I find that institutions of private property are helpful in explaining differences in growth between two countries at the border, while contracting institutions, local and national levels of public goods, as well as education and cultural variables, are not. The last chapter of my thesis, which I have published in the *Journal of Public Economics*, investigates the dynamics of the world distribution of income using more robust methods than those in the previous literature. I derive sharp bounds on the Atkinson inequality index for a country's income distribution that are valid for any underlying distribution of income conditional on given fractile shares and Gini coefficient. I apply these bounds to calculate the envelope of possible time paths for global inequality and welfare in the last 40 years. While the bounds are too wide to reject the hypothesis that world inequality may have risen, I show that world welfare rose unambiguously between 1970 and 2006. This conclusion is valid for alternative methods of dealing with countries and years with missing surveys, alternative survey harmonization procedures, alternative GDP series, or if the inequality surveys used systematically underreport the income of the very rich, or suffer from nonresponse bias.

Thesis Supervisor: Daron Acemoglu
Title: Elizabeth and James Killian Professor of Economics

Thesis Supervisor: Jerry Hausman
Title: John and Jennie S. MacDonald Professor of Economics

Contents

Acknowledgements

The greatest and deepest debt that I owe is to my grandmother, Edda Resina, and to my parents, Lev Pinkovskiy and Izabella Pinkovskaya, for sharing every second of my existence with me, including these five years of writing my thesis. For the duration of my doctoral studies, my grandmother moved with me to Massachusetts and lived with me, despite advanced age, indifferent health, a complete change in lifestyle, and the rigors of frequent travel, so that we may support each other through this time and remain as close as we have always been. She has spoken the truth even when I wasn't inclined to listen, and taught me that one cannot go far in the carriage of the past, and that one must always be wary of confusing that which glitters with gold. My mother visited us almost every week and my father came whenever his intensive work permitted. I am overjoyed now to return home, to give you as much love as I have within me, and to grow to be able to give that kind of love to the family I shall have in the future. Thanks to you, I have never felt alone despite the many difficulties and disappointments on the way, and I always had someone with whom to share the many moments of joy and discovery. Thanks to you, I have never felt my work to be without meaning, or my efforts to be without end. To you I dedicate this dissertation, as I dedicate every action of my life that may be worthy of you.

I also must honor the memory of my grandfather, Simon Isayevich Pinkovskiy, who watched over me from Russia, our native land, and who kept abreast of all my developments and was able to rejoice at their completion, though, sadly, not at their ceremonial conclusion. Thank you for giving me what you gave me when I was a child and throughout.

My next debt is to my dissertation advisers, Daron Acemoglu, Jerry Hausman and Amy Finkelstein. Thank you so much, Professor Acemoglu, for giving me questions worthy of spending a lifetime on, for tirelessly guiding me towards the best practices of economics, and for your indefatigable patience, even through my occasional obstinacy. Professor Hausman, I will be eternally indebted to you for your faith and confidence in me since the day I stepped over the threshold of E-52, for teaching me so many new tools and methods of extracting inferences from data, for your unparalleled kindness to me as your teaching and research assistant, and for your unflagging encouragement that has many times made the difference between abject failure and a successful project. Finally, thank you so much, Professor Finkelstein, for showing the graph that ultimately inspired the main paper of my dissertation, and for teaching me so much about empirical work.

I also must thank my professors and collaborators at MIT, Harvard and Columbia University, who have set me on the road of research. Xavier Sala-i-Martin welcomed me as a college sophomore into his projects, and I have felt his faith in me and support for me ever since, whether I was at Columbia or MIT. Now, we have produced many joint papers, and I look forward to many more. James Snyder was the first at MIT to take me on as a coauthor, and continued watching over me and giving valuable suggestions even after moving across town to Harvard. Shigeo Hirano and Gabe Lenz, Jim's coauthors on the project in which I participated, taught me a great deal about looking at and interpreting data. Pierre-Andre Chiappori introduced me to the fascinating world of social insurance in health care and our collaboration has been vital to launching me on the main project of my dissertation. Bernard Salanie read many of my drafts and gave me comments almost as though I had been his student. Jon Gruber, Heidi Williams, Ben Olken, Elias Papaioannou, Daniel Posner, Jens Hainmueller, Anna Mikusheva, Victor Chernozhukov, Whitney Newey and Emmanuel Saez offered vitally needed advice and suggestions on where to seek out data and what methods to use throughout my time at MIT and I hope to put their lessons to good use in the future. Finally, I would like to thank my sagacious Columbia advisers Michael Woodford and Susan Elmes, and to honor the memory of Leonardo Bartolini, who gave me my first job.

Throughout my time at MIT, I relied immensely on the kindness, patience and boundless good will of Gary King, Peter Hoagland, Beata Shuster, Patty Glidden, Emily Gallagher, Loida Morales, Kara Nemergut, Cherisse Haakonsen, Lauren Fahey and Kim Scantlebury. Thank you so much for always being there to help, even when the task in question was my responsibility. Without you, teaching, research and especially the job market would have come to a standstill for me, and for all of MIT.

I have benefited from great support from my friends and colleagues at MIT. To Adam Sacarny and Horacio Larreguy I owe many thought-provoking conversations on health and political economy, and equally importantly, immense sympathy, encouragement, and willingness to devote valuable time to me and give thoughtful and sincere feedback. Thank you for watching my back in sickness and in health, and for being the first to listen and the last to be silent. Pablo Querubin, David Chan, Greg Leiserson, Christopher Palmer, Tyler Williams and especially Joe Shapiro gave me very valuable comments on my main paper when I needed them most. Melanie Wasserman, Luu Nguyen, Brendan Price, Ariel Zucker and Rachael Meager shared with me many thought-provoking conversations. To Vincent Pons, Johann Frick, Ekedi Mpondo-Dika, Alex Peysakhovich and Aurelie Ouss I owe many happy evenings of exceptional and multilingual conversations. Finally, I would like to thank Sarah Venables and Jennifer Peck for their kindness, warmth and hospitality, and Xiao Yu Wang for moral support and for causing our whole class to miss out on the loneliness associated with doctoral studies.

I would also like to acknowledge my debt to my many friends from outside MIT, who have so often reached out to me and sustained me especially when the going got rough. Thank you so much, Morgan Hardy, for your many trips to Cambridge, on your own or with Tim, Glancy, Judith and Cleo, for our beautiful walks and philosophical conversations, and for your tireless spunk. Shuky BarLev-Ehrenberg, thank you for being like an older brother to me and *mazel tov* to your growing family. Alex Dean, though we tended to speak from afar in space and time-zones, you have been such a good friend to me, as you have been since we first met in middle school. I am also very thankful to John Klopfer, Kate Zhang, Leora Kelman, Kim Nguyen and Weizhong Ji for keeping the spirit of Knickerbocker alive in Cambridge and Concord. To Joshua Feigenbaum and Susan Jaffe I am grateful for their everpresent interest in my work. Drs. Yuri and Ninel Zamdborg provided first-person evidence on practicing medicine during the managed care backlash, and Natalya Shnitser gave me an excellent account of the backlash's legal aspects. Finally, I am very grateful to my cousin, Luba Nisenbaum, and to Isaac Gorodetski, Dmytro Karabash, Mikhail Shklyar and Diana Dreyer for extremely needed and valued moral support at a particularly intense time.

Over the course of these five years, I have received financial and intellectual support from a variety of organizations that have helped me concentrate on my research rather than on quotidian challenges, and that have inspired me to new and better questions. I am grateful to the Paul and Daisy Soros Fellowship for New Americans for intellectual stimulation and for including me into its circle, to the NSF Graduate Research Fellowship Program for funding, and to the Institute of Humane Studies for the Humane Studies Fellowship and for inviting me to so many thought-provoking seminars.

Finally, I would like to thank the students who have had the responsibility of enduring me as their TA during my time at MIT. There are too many of you to list, but I would like to highlight Yangzhou Hu, Ndubisi Onuora, Qian Lin, Noam Angrist, Matt Curtis, Andrew Stuntz, CJ Enloe, Yaroslav Mukhin, Rachael Meager, Matt Lowe, Peter Hull, Nick Hagerty, Elizabeth Setren, Stacy Carlson and Lauren Linzemeyer. It was your enthusiasm and willingness to try new things that has given me encouragement to continue on.

Chapter 1

The Impact of the Managed Care Backlash on Health Care Costs: Evidence from State Regulation of Managed Care Cost Containment Practices

1.1 Introduction

Controlling health care costs is a major unmet challenge for public and private health care systems in the United States. Personal health care spending has nearly doubled as a share of GDP in thirty years, rising from 8% of GDP in 1980 to 14.8% of GDP in 2009, and often has grown at a linear (and hence, unsustainable) rate for decades at a time. The problem of stemming the growth in health care costs is particularly urgent because such costs form a significant part of U.S. government spending, particularly with the passage of the Affordable Care Act (ACA), which will lead to substantial government subsidies to individuals to purchase private insurance.

The overall trend of rising health care costs in the U.S. saw a temporary break during the 1990s, when personal health care spending as a share of GDP remained nearly constant (actually, declined slightly) from 12.1% in 1993 to 11.94% in 2000. This stabilization of health care costs coincided with the peak of the so-called managed care revolution, which saw the replacement of conventional insurers (who reimbursed hospitals and physicians for services provided without regulating utilization) by health insurance organizations that managed the medical care of their enrollees. The organizational innovation of managed care firms was to integrate physicians and insurers partially or completely to align their incentives and discourage physicians from inducing demand for medical care. The most well-known type of managed care organization, the HMO, restricted its patients to see a strictly delimited network of providers, who sometimes were its employees. While the growth of health insurance premiums slowed significantly, patients and physicians chafed under managed care controls. At the end of the 1990s, there arose a widespread backlash against managed care cost containment practices, with increasingly negative media coverage of managed care. Ultimately, state governments passed "patients' bills of rights" that limited the ability of managed care firms to restrict care and shape the incentives of medical practitioners. Health care costs resumed rising as a share of GDP in 2001, at the height of the managed care backlash. It remains an open question whether managed care succeeded in stabilizing U.S. health care costs or whether the slowdown in U.S. health care cost growth in the 1990s was a product of other factors (Glied 2003).

This paper will find that the managed care backlash, as proxied by the amount of legislation passed to restrict managed care cost containment practices (hereafter, backlash regulations), in fact had a causal effect on increases in health care costs. My identifying assumption is that backlash regulations increased health care costs only to the extent that managed care was already containing costs in the given state, while the timing of backlash regulations is exogenous with respect to all other variables whose effect on changes in health care costs is a function of managed care intensity. This assumption is weaker than the standard difference-in-difference assumption that the timing of the backlash regulations is uncorrelated with shocks to health care costs. My assumption is plausible because backlash regulations are politically determined variables, which are likely to arise from distinct data generating processes than are outcomes in health care markets. However, it could fail in various ways: for instance, if regulations are passed in response to severe cost containment, which also decreases health care share, or if regulations are correlated with other trending variables in the health care market. Robustness checks involving more sophisticated control variables, a dynamic analysis of the passage of regulations and an instrumental variables analysis help rule out such concerns.

To obtain my findings, I use panel variation in the passage of backlash regulations, which were passed in different years and in different numbers in different states. I include both the main effect of backlash regulations as well as, crucially, its interaction with managed care intensity. I proxy managed care intensity by HMO penetration in each state in 1995. HMO penetration is a natural proxy for managed care intensity both, directly, because HMOs are the most restrictive form of managed care, and, indirectly, because looser managed care organizations in the same state had to cut costs more substantially to compete with the HMOs.[1] Furthermore, I explicitly model the substantial persistence in the health care share by estimating models with the lagged health share as a regressor. An econometric difficulty in estimating such models is their mechanical failure of strict exogeneity and the poor performance of instrumental variables estimators when the persistence of the dependent variable is high (as documented by Hausman, Hall and Kuersteiner 2007). Therefore, I use a novel approach pioneered by Hausman and Pinkovskiy (2013) that avoids the bias of instrumental variables by estimating a transformed version of the lagged dependent variable model with fixed effects via nonlinear least squares.

My results indicate that because of the managed care backlash, health care costs in a state with average HMO penetration in 1995 grew by 0.1 percentage point more per year than they would have otherwise, which is equal to the average change in the health care share across states in 2005. To assess the magnitude of my result, I use my regression to make a dynamic counterfactual forecast of the evolution of each state's health care share under the assumption that the number of backlash regulations was equal to zero in every state and year, and aggregate the forecasts to predict the counterfactual for the U.S. health care share for each specification I run. I find that under the counterfactual of no managed care backlash, the U.S. health care share in 2005 would have been 12.3%, more than a full percentage point of GDP lower than the actually observed level, and close to the 1993 level of 12.1%. In fact, most of the rise in the U.S. personal health care share of GDP between 1993 and 2005 (from 12.1% to 13.5% of GDP), as well as the fact that health spending resumed rising as a share of GDP after remaining stable at 12% between 1993

[1] The performance and prevalence of HMOs also could have provided demonstrations to less managed health care plans that tightly managed policies are marketable, encouraging these plans to adopt them. I provide evidence that HMO penetration is correlated with tight management practices (the degree of restriction on patients seeing providers) in Section 3.

and 2000, can be statistically explained by backlash regulations. I provide a variety of robustness checks for my identifying assumption by including state trends and covariates, accounting for the timing of the passage of the regulations, varying the geographic unit of analysis, accounting for other health insurance regulations being passed at the time and employing instrumental variables. To my knowledge, this is the first paper quantifying the effects of the managed care backlash on health care costs. Glied (2003) considers the reasons for the resumption of health care cost growth in the early 2000s, but does not give a quantitative estimate for the possible effect of the backlash. A large literature in health policy and law (Peterson 1999) has studied the managed care backlash qualitatively, discussing the reaction of the public, the legislation passed, and the weakening of managed care cost containment practices, but has not calculated the impact on health care costs.

I do not attribute the health care cost rise I find to the direct effect of backlash regulations specifically, but to the managed care backlash in general. The backlash regulations may have been unevenly enforced across different states, and may have acted as warnings to insurance companies rather than as binding constraints on their practices. However, I do provide evidence by analyzing the timing of the passage of the backlash regulations and by including smoothly varying state trends and covariates that the backlash regulations do not reflect changes in preferences that may be correlated with shocks to health care and health insurance markets.

It is interesting to examine the mechanisms by which the managed care backlash may have increased health care costs. The literature on managed care has investigated the effects of managed care on provider salaries (Cutler, McClellan and Newhouse 2000) and length of stay (Glied 2000). Using the same methodology as for my baseline result on health expenditures, I look at the effects of the managed care backlash on these health care inputs. I find evidence that the managed care backlash raised the salaries of medical providers. I also find that the data is consistent with the hypothesis that the managed care backlash increased lengths of stay.[2] Finally, I consider whether the managed care backlash is associated with health improvements. I examine the effects of the managed care backlash on mortality because other health outcomes exhibit composition bias and are therefore difficult to interpret. I find that backlash regulations (and hence, the associated health care cost increases) are not associated with strong and unambiguous decreases in mortality, but the confidence intervals of my estimates are wide.

The rest of the paper is organized as follows. Section 1.2 presents a brief history of managed care in the United States and describes managed care cost containment practices as well as the laws regulating them. Section 1.3 describes the data. Section 1.4 explains the empirical specification. Section 1.5 presents the baseline results for health spending growth, as well as the associated robustness checks. Section 1.6 presents results for health resources utilization and mortality. Section 1.7 discusses political determinants of backlash regulations and presents an instrumental variables analysis. Section 1.8 concludes.

1.2 Institutional Background and History

Since patients and doctors have substantial flexibility in choosing the intensity of treatment, the health insurance market suffers from moral hazard (Arrow 1963) unless insurers monitor treatment choices or use financial incentives for insurees to economize on care. Most U.S. health insurance before the 1980s (and all of Medicare and Medicaid) was conventional: insurers reimbursed physicians

[2] In addition, I look at the effects of the managed care backlash on technological intensity (Cutler and Sheiner 1998; Baker and Wheeler 2000, Baker and Phibbs 2002), as measured by the number of facilities in hospitals per capita. While my estimates are noisy, they are consistent with the number of facilities in hospitals per capita increasing with the managed care backlash.

and hospitals for each procedure performed, using deductibles and copayments to provide incentives against unlimited utilization, but they did not intervene in physician treatment choices. An alternative arrangement, referred to as managed care, involves insurers directly contracting with or even employing physicians and regulating their choice of care either through more sophisticated financial incentives or through the threat of termination (or "deselection" from the contract network) if the insurer deems that the physician utilizes health resources beyond what is clinically necessary. The most restrictive variety of managed care, the health maintenance organization (HMO) either hires the physicians whose care it reimburses, or forms exclusive contracts with a panel of physicians, forbidding its patients to see other physicians in most circumstances. A less restrictive (and currently most widespread) version of managed care is the preferred provider organization (PPO), which contracts with a network of physicians to receive discounts on their fees in return for the PPO giving a discount to its patients to see the physicians in the network. HMOs and POSs depart from fee-for-service reimbursement by paying physicians salaries, bonuses for low utilization, or capitated reimbursement for each patient regardless of cost. Additionally, managed care firms restrict patient choices through gatekeeping (the requirement to see specialists only after a referral by a primary physician) and utilization review (submission of proposed procedures to the insurer, and potential refusal to cover expensive or experimental treatments).

Contracts between insurers and physicians for prepaid health care have been observed as early as the 19th century, and the roots of the HMO Kaiser Permanente trace back to the 1930s. However, for most of the 20th century, managed care remained a small fraction of the U.S. health insurance market because of state regulations denying hospital privileges to managed care-employed doctors and preventing managed care from advertising, which were repealed over the course of the 1970s (Feldstein 1988). As health care costs continued to rise during the 1980s, more and more employers and individuals saw relatively less expensive managed care as preferable to conventional fee-for-service insurance.[3] As late as 1989, 73% of privately insured Americans had conventional insurance; by 1996, this fraction was just 27%, and by 2005, conventional fee-for-service insurance covered only 3% of the privately insured market (Kaiser Family Foundation 2011). The fraction of people in HMOs rose from 5% of the total insured population in 1980 to 30% in 1998. At the same time, the health care cost share in the United States, which was trending upward throughout the 1980s, leveled off and even slightly decreased during the 1990s.[4]

To the extent that they lowered the level and growth rate of medical costs and insurance premiums, the cost containment practices of managed care benefited healthy patients, employers and the federal and state governments. However, they hurt physicians, who now had to compete for membership in the networks of managed care organizations and incorporate financial considerations of the cost of treatment into their practice style, as well as less healthy consumers, who now obtained much lower quality insurance. The employment-based system of health insurance served to increase the salience of discontents with managed care and decrease the salience of their advantages because the wage increases resulting from cheaper health insurance were not explicitly tied to the change in health insurance arrangements in the minds of workers (Blendon et al. 1999).[5] Instead, workers suffered the disruption of switching not only to a new insurance regime but also to a new provider network without attributing any of the resulting wage increases to the switch to managed care.

[3] Another reason for the rapid spread of managed care may have been a series of insurance market reforms for the small group (small business) market. (Buchmueller and DiNardo (2002), Buchmueller and Liu (2003)).

[4] Managed care was integrated into Medicare through the voluntary program Medicare part C, which allowed patients to opt out of traditional Medicare in favor of a managed care plan. Medicaid employed managed care by shifting patients to it by fiat at the state (or sub-state) level.

[5] A robust finding in the health care labor literature is that increases in health insurance premiums are shifted almost completely to worker wages. See e.g. Gruber (1994, 1997).

As a consequence of these discontents, in the late 1990s a powerful cultural, media and legal backlash took place against managed care in general and HMOs in particular. HMOs were depicted in special reports in major newspapers and in popular films such as *As Good as it Gets* as impersonal, greedy bureaucracies that denied life-saving care to critically ill people in order to enhance their profits. Brodie et al. (1998) document that the tone of media coverage of managed care, especially in the most visible news sources such as television and newspaper special reports, grew to be increasingly critical, and gave increasing weight to anecdotes of managed care patients being denied essential care. Partially in response to this backlash, legislation was initiated at both the state and the federal level to create "patients' bills of rights" that would limit the cost-control practices that managed care organizations would be allowed to use. The backlash regulations took four important forms as documented in Table 1.1 : regulations to provide access to physicians and treatments, regulations to provide venues for appealing managed care denials of coverage, regulations of the insurer-provider relationship and regulations to mandate particular procedures. Access regulations took the form of permitting patients to see doctors outside a managed care firm's network if these doctors had previously been treating the patient for a long-term illness (continuity of care), allowing patients to have direct access to specialists without having to first go through a primary care gatekeeper physician, and mandating coverage of emergency room use regardless of whether a sufficiently severe health problem was actually uncovered. Appeals regulation required managed care firms to create credible procedures for reviewing its coverage decisions, either internally or relying on an outside arbitrator. In several states, managed care firms could be sued for medically adverse events resulting from denial of coverage. Provider regulation limited the ways in which managed care firms could reimburse physicians in their networks or in their employment, limited managed care's control over the composition of their network (Any Willing Provider or Freedom of Choice laws), and forbade managed care firms to prevent their providers from disclosing information to patients about treatments not covered by the insurers (gag clauses). Mandated benefits were mostly focused on maternity stays, reconstructive surgery after cancer, and diabetes supplies. While no federal legislation was passed[6], nearly all states passed various legislation of their own, at different times and of differing severity, but almost all in the 1995-2001 period. By 2001, health care costs and insurance premiums had resumed rising, while HMOs were being displaced by the less fiscally stringent PPOs. The percentage of people insured by HMOs, once at about 30%, declined to 23%.

1.3 Data

1.3.1 HMO Penetration

I obtain data on HMO penetration indirectly from the survey firm Interstudy. I obtain state-level data for the percentage of the total population (including Medicare and Medicaid recipients) enrolled in HMOs for 1980, 1985, 1990 and 1995-2007 from the Statistical Abstract of the United States. I use HMO enrollment, rather than total enrollment in managed care, to measure managed care intensity because by the beginning of my sample period (1995-2005), most U.S. private health insurance was some form of managed care, with HMOs being the most restrictive, while the share of conventional insurance was low and falling, and thus, unlikely to be very informative. I also obtain data on total population HMO penetration at the county level for the years 1990-2003 from

[6]H.R. 2723, sponsored by Democrat John Dingell, passed the house, but the Senate version of the bill was so mild that Representative Dingell urged the Democrats in the House to vote against it in the joint session.

Laurence Baker, who constructed these measures using unit records from Interstudy, which are not available for the public. The exact method of construction of the county-level data is described in Baker and Phibbs (2002) and involves extrapolation on the basis of county population and the regional enrollment of HMOs serving each county in question as reported to Interstudy. The Statistical Abstract and Laurence Baker use somewhat different definitions of HMOs to construct their HMO penetration rates, but my results are robust to using either measure in the state-level analysis. For regressions at the state level I use the Statistical Abstract series, and for substate-level regressions, I use the Baker series.

Throughout this paper, I use HMO penetration as a proxy for the intensity of managed care activity in a given region (state, MSA, county). It is intuitive that HMO penetration should be a good proxy for the overall level of managed care activity because HMOs were the most restrictive form of managed care. Since HMOs and less restrictive forms of health insurance operate in the same product and factor markets, high HMO penetration should incentivize other insurers to adopt restrictive practices to lower costs so that they could better compete with HMOs. The presence of HMOs should spread restrictive cost containment practices through the "demonstration effect" of showing that the health insurance market will bear such practices (e.g. that large numbers of people will purchase plans that do not cover all local providers). As discussed by Bloch and Studdert (2004), physicians and hospitals would be likely to use the same practice style for all their privately insured patients, whether those belonging to HMOs or not, which would lead to spillovers. A large literature in managed care documents that premium growth rates within and outside HMOs track each other very closely (Ginsburg and Pickreign (1996, 1997) use KPMG data to show that HMO premium growth was at least 75% of conventional premium growth over the period 1992-1996), and a series of papers shows that increases in HMO penetration in a region decrease the health cost growth rate of conventional insurers in the same region (Baker 1997, Chernew et al. 2008). HMO penetration also correlates very well with evidence of restrictive cost containment practices. The MEPS-IC, which is a nationally representative survey of health insurance plans, asks about the extent to which a plan contracts selectively, and about the extent to which care is managed in the plan, with answers to these questions being independent of whether a plan is formally an HMO (so a conventional plan without selective contracting but with some utilization review would answer "yes" to the question of whether there is any managed care in the plan). Figure 1.1 shows the correlation between HMO penetration in a state and the state-level estimates of the number of firms that offer plans with any managed care from the 1996 MEPS-IC (Correlations between HMO penetration and the extent of exclusivity of providers are even stronger).[7] We see that the correlation is tight, which reinforces our confidence in HMO penetration as a proxy for the intensity of managed care cost containment practices.[8]

Figure 1.2 shows a time plot of the Statistical Abstract HMO penetration measure for the United States as a whole. We see the steady rise of managed care during the 1980s and the 1990s, followed by a partial but precipitous decline during the backlash period. Figure 1.3 shows a map of (Statistical Abstract) HMO penetration by state in 1995. We see that HMOs were strongly clustered regionally, with high penetration on the West Coast, in the Northeast (especially Massachusetts)

[7] The 1996 MEPS-IC was not large enough to support state-level estimates for 10 of the smallest states; hence, this correlation is on the basis of the 40 largest states only.

[8] I prefer the HMO penetration measure to the MEPS-IC measures because the MEPS-IC statistics are liable to have measurement error. MEPS-IC publishes statistics only on the fraction of firms offering plans with various levels of intensity of managed care, rather than on the number of people enrolled in any such plans. Since large firms tend to have different health insurance purchasing behavior than do smaller firms, I do not expect the two measures to be the same. Moreover, since health care costs depend on the number of patients involved rather than on the number of firms involved, I prefer the population-based HMO penetration measure to the firm-based MEPS-IC measures.

and in the Midwest.

1.3.2 Backlash Regulations

The key independent variable in my analysis is state regulation of managed care cost containment practices. I obtain data on the passage of various managed care regulations during the backlash from the National Council of State Legislatures, which maintains databases of state laws on various topics for research purposes freely available to the public. Each type of regulation is listed separately for each state, even if multiple regulations were passed together in a single bill, and multiple regulations on a single topic (e.g. banning financial incentives for physicians) are listed separately. Altogether, there are about 750 backlash regulations. Table 1.1 shows the different types of regulations, both in a fine (27 groups) and in a coarse (4 groups) categorization, as well as how many regulations of each type were passed. Figure 1.4 shows a time series of the adoption of new backlash regulations. We see that most such regulations were passed in the 1996-2001 period, although a few were passed before and after this period. No new backlash regulations were passed after 2005. In my analysis, I will use the raw total of backlash regulations as a measure of regulation intensity in most specifications, although I will check for robustness to alternative parametrizations of the regulations.

Throughout the paper, I use backlash regulations as a proxy for the intensity of the backlash in general, and I do not assert that the effect I find is the causal effect of the regulations themselves. The heterogeneous nature of the regulations and of their enforcement precludes such a causal attribution. Moreover, the passage of regulations may have signaled to managed care organizations that more binding legislation may be passed if they do not change their practices. Nevertheless, it is important to determine that I am not attributing the effects of some other policy in the health care sector to the managed care backlash, which I do in Section 1.5. [9]

I remain agnostic whether the backlash regulations limited the existing levels of managed care cost containment, or whether they limited the future growth of managed care cost containment practices, both in geographical scope and in intensity. In particular, backlash regulations may not have been effective in controlling how managed care firms regulated access to existing medical technologies, but they may have been effective in making managed care firms allow subsequent medical technologies. If technological progress in medicine drives health care costs, then such an effect pattern could generate substantial increases in health care costs over time without visibly reversing any aspect of managed care policy.

We see that backlash regulations are not associated with pre-period state HMO penetration. Figure 1.5 shows a map of the regulations in 2005, and Figure 1.6 presents a scatterplot of the number of backlash regulations passed by 2005 against HMO penetration in 1995. The relation is positive, but weak and insignificant.[10] We see that some states with low HMO penetration (like Wyoming and Mississippi) also had few regulations. However, some states with low to moderate HMO penetration (Texas, South Dakota, Virginia, Kentucky, Tennessee) were leaders in backlash regulations, while the managed care leaders (California, Oregon, Massachusetts) had lower levels of backlash regulations. Moreover, aside from Kentucky and Tennessee, there does not seem to

[9]The incidence of backlash regulations is not straightforward because public insurance (Medicare and Medicaid) tended to be regulated separately from private insurance, and because self-insured firms were exempt from state regulation through ERISA. As I have discussed in Section 1 and earlier in Section 3, there is good reason to believe that backlash regulations had substantial spillovers to insurers who were not regulated by them directly because of the extent of spillovers between HMO and non-HMO insurance.

[10]A 10 percentage point increase in 1995 HMO penetration is associated with an additional 0.6 regulations, with a t-statistic of 0.6.

be much geographic clustering of backlash regulations. I explore controlling for other potential time-varying correlates of backlash regulations in Section 1.5.

1.3.3 Other Regulations

I obtain data on other health insurance regulations from the Blue Cross Blue Shield publication "State Legislative Health Care and Insurance Issues," which was sent to me directly by its author, Susan Laudicina. From this data, I extract the series of state mandated benefits, the series of state small-group insurance reforms, and the series of state individual insurance reforms. Since mandated benefits are qualitatively similar (although involving mandates of different expense), I use the raw total number of mandated benefits in each state-year as an independent variable. However, since different small-group and individual insurance reforms regulate different aspects of the insurer-insuree relationship, I follow Simon (2005) and code whether each state has a "full reform" or does not have a "full reform." I define a full reform by the presence of a guaranteed issue law, a guaranteed renewal law, and rating reform.

Another important set of health insurance reforms taking place in the late 1990s was Medicaid expansion. The observed fraction of people in a given state who are eligible for Medicaid is endogenous, because it depends on state economic conditions and demographics, all of which may affect the health share of GSP. Currie and Gruber (1996) create a time series of "simulated Medicaid eligibility," which is the fraction of people in a standard population who are eligible for Medicaid according to the laws of a given state. This series has most recently been updated by Gruber and Simon (2008). I obtain the simulated eligibility series updated to 2004 directly from Kosali Simon.

1.3.4 Dependent Variables

I obtain state-level data on economic activity (gross state product) and data on total (public and private) personal health expenditures as well as separate data on personal health expenditures in Medicare and Medicaid from the National Health Expenditure Accounts, maintained by the Center for Medicare and Medicaid Services (CMS). I also obtain county-level data on economic activity (personal income) from the Bureau of Economic Analysis, which I use to normalize my health spending variable when I run regressions at sub-state levels. To obtain data on health expenditure at the substate level, I use the American Hospital Association Annual Survey, which provides disaggregated data on hospital expenditures. I also use the AHA Annual Survey to obtain state-level data on hospital payrolls, employment, admissions, utilization and technology choice (number of facilities of various types). I also obtain state-level data on employment and salaries in the ambulatory health sector, which comprises of physician offices, outpatient centers and home health care, from the Bureau of Economic Analysis. Unfortunately, I do not have data on total physician expenditures at a substate level, and neither do I have hospital-level data on expenditures reimbursed by Medicaid and Medicare, so in my sub-state analysis, I am restricted to analyzing total hospital expenditures. Finally, I obtain data on mortality rates by state and year from the Center for Disease Control.

Table 1.2 presents summary statistics for state-level data in 2005, including personal health expenditures, regulations, and HMO penetration. We see the sample mean of backlash regulations in the entire dataset was about 15, and the sample mean of 1995 HMO penetration is 14.5%. The mean annual change in the health care share of GSP in a typical state was about 0.1 percentage points.

1.4 Empirical Strategy

It is intuitive that health spending is very persistent. The set of sick and healthy people, their medical needs, and the practice styles and technology used to treat them tend to be the same over time, because of the relatively unchanging landscape of human illness and because rapid change in the medical system would be unsettling to patients. The persistence of health spending is found to be important in papers in which it is modeled, such as Cutler and Sheiner (1998). Furthermore, many papers find that institutional changes in health care markets have effects not only on the level, but on the trend (or the growth rate) of health care spending or of utilization patterns in the health care sector (Finkelstein 2007; Acemoglu and Finkelstein 2008). I therefore estimate a flexible dynamic panel specification that allows the lagged value of health care costs to affect the current value of health care costs. The specification I estimate is:

$$P_{s,t} = \alpha_s + \lambda_t + \delta P_{s,t-1} + \beta R_{s,t-1} + \gamma R_{s,t-1} \times HMO_s^{1995} + X'_{s,t}\eta + \varepsilon_{s,t} \qquad (1.1)$$

where $P_{s,t}$ is the total health spending share of gross state product in state s and year t (in other regressions, the dependent variable will be different), α_s and λ_t are state and year fixed effects respectively, $R_{s,t-1}$ is the number of regulations in force in state s in year $t-1$, HMO_s^{1995} is HMO penetration in state s in 1995, and $X_{s,t}$ is a vector of controls (absent in the baseline specification). The coefficients of interest are γ, the interaction effect of regulations on health spending as a share of GSP as a function of HMO penetration, β, the level effect of regulations as a function of HMO penetration, and the persistence parameter δ.

My identification assumption is that states with different pre-period HMO penetration have differential trends in health care costs as a share of output in the period 1995-2005 only because of backlash legislation, taking into account the natural persistence of the health care cost share of GSP. In particular, because I use panel data with fixed effects, I avoid the potential danger that states with different amounts of regulation also differ in other static characteristics that influence health care costs as a share of GSP.

It is well known (Anderson and Hsiao 1982; Arellano and Bond 1991; Blundell and Bond 1995) that estimation of equation (1.1) by ordinary least squares yields biased and inconsistent estimates of the coefficients δ, β and γ. The standard technique for dynamic panel estimation is the approach of Arellano and Bond (1991) of differencing equation (1.1) and using lagged dependent and independent variables as instruments for the lagged difference via GMM. However, this approach exhibits substantial bias in the case when δ is close to unity because the correlation between the instruments and the endogenous variables is close to zero (Hahn, Hausman and Kuersteiner 2007). In particular, the coefficient δ tends to be biased downward, suggesting less persistence in the dependent variable than is actually present. Therefore, in this paper, I follow Hausman and Pinkovskiy (2013) and estimate equation (1.1) by back-substituting for $P_{s,t-1}$ to express $P_{s,t}$ in terms of $P_{s,0}$ and lags and levels of the independent variables, and estimating the resulting equation by nonlinear least squares. I provide a complete description of the procedure I use in Appendix II, as well as several additional tests for the exogeneity of $P_{s,0}$ and for the robustness of the results if the regressors are predetermined rather than strictly exogenous. In Section 1.5, I show a version of my baseline estimates computed using the Arellano-Bond method, and note that all the coefficients are lower than using the nonlinear least squares method, as predicted by Hahn, Hausman and Kuersteiner (2007) given the size of the parameter δ.

1.5 Results: Cost Growth

To assess the magnitudes of my estimates, in all my tables, I present forecast values of the total health spending share of U.S. GDP (or the Medicare, Medicaid or private share in some specifications) under the assumption that no backlash regulations had been passed. Forecasts are obtained by bootstrapping the coefficients on the terms in the model that depend on backlash regulations (and on the lagged dependent variable if it is present) and computing the increase in the dependent variable coming from backlash regulations for each state and year. I then subtract the bootstrap estimates from the true values of the dependent variable (in levels) for each state and year, and aggregate the state-level forecasts (with suitable weights) to obtain a national forecast. I repeat this procedure 500 times, each time drawing a different set of coefficients from the estimated distribution. Since the dynamic forecasts involve powers and products of correlated normal random variables, the resulting forecast distribution is non-normal, and in particular, severely right-skewed. [11] Therefore, I report the median of the resulting bootstrap forecasts. The upper and lower 95% confidence bounds of the forecast distribution are reported below the point forecast.

1.5.1 Baseline Results

Table 1.3 presents estimates of equation (1.1) when the dependent variable is the total personal health spending share of GSP, the private share, the Medicare share and the Medicaid share. We see that the coefficient of interest – the coefficient on the interaction between backlash regulations and pre-period HMO penetration – is significant when the dependent variable is the total share or the private share. The magnitude of the interaction coefficients when the dependent variable is the total health share is 0.093 percentage points, and is comparably high for the private share. The (insignificant) main effect of regulations is (-0.007) for the total share, and similarly for the private share. Since the average number of regulations in 2005 is 15.22, and the average 1995 HMO penetration is 0.145, for a typical state, the managed care backlash is associated with an extra 0.1 percentage point increase in the personal health share of GSP every year.[12] Given that the mean increase in the personal health share of GSP across all states in 2005 was also 0.1, we see that the estimated effect of the backlash is substantial. The interaction coefficient when private share is the dependent variable is slightly smaller than the interaction coefficient when total share is the dependent variable, but the private health share of GDP is approximately two-thirds of the total health share of GDP, so the proportional effect of the managed care backlash on the private share is larger than on the total share.

The counterfactual predictions of the model for what would have happened without the managed care backlash are striking. The total health share of U.S. GDP was 13.48% in 2005, but without backlash regulations, it would have been 12.28%, about 1.2 percentage points of GDP lower, which, given that U.S. GDP in 2005 was about 12 trillion dollars, amounts to 144 billion dollars lower. This is equal to 46% of Medicare spending in 2005 (which was 2.6% of GDP) and is 10% of the counterfactual health care share in 2005. The confidence interval of this forecast, however, is very large, and does not permit us to rule out the observed 2005 level. Figure 1.8 plots the observed path of the total share of GDP and its counterfactual under the assumption of no regulations; we see that without backlash regulations, the model suggests that the total health care share of GDP would

[11] I present an estimate of the baseline forecast distribution in Figure 1.7.

[12] We have $15.22 * (0.093 * .145 - 0.007) \approx 0.1$

have tended to be close to 12%, its long-run level during the 1990s. [13] A similarly low forecast, also insignificantly different from the observed 2005 level, can be observed for the private share.[14] Since the point estimate of the lagged dependent variable coefficient is less than unity (it is 0.93), we can obtain a cumulative effect of the managed care backlash. As mentioned before, the one-year increase in the health care share of GSP associated with backlash regulations is 0.1 percentage point, so the long-run impact will be given by $\frac{0.1}{1-0.93} = 1.43$ percentage points. This is close to the rise in the U.S. health care share of GDP that we observe between 1999 and 2005, suggesting that the additional health care share growth from the managed care backlash may have exhausted itself by 2005. However, since the upper bound of the lagged dependent variable coefficient is close to unity, much larger long-run impacts of the managed care backlash are consistent with the data.

Since HMOs accounted for only 30% of the insured population at the height of the backlash, it is obvious that much of the effect of the managed care backlash was a spillover effect to non-HMO insurance (conventional and looser managed care arrangements) rather than a direct effect on HMOs. As discussed in Section 1.3, such spillovers are both theoretically expected and empirically documented in the managed care literature. Some channels for this spillover will be shown in Table 1.9, where we will see that backlash regulations are associated with increases in hospital salaries and, possibly, with length of stay and the number of hospital facilities, which should have impacted hospital spending beyond that on HMO patients.

The associations between backlash regulations and the Medicare and Medicaid shares of GDP are different in magnitude and nature from the effects on the total and on the private share. We see that the counterfactual estimates for Medicare and Medicaid both show cost increases that are smaller relative to those in private insurance. We also see that the main effect and interaction coefficients are insignificant and relatively small (and the interaction coefficient when the Medicaid share is the dependent variable is negative). One rationalization of these results is that Medicare is a federal program with a federal-level reimbursement schedule that creates high-powered incentives (Clemens and Gottlieb 2012) and should therefore not have been directly affected by backlash regulations. Medicaid, though regulated by the states, has its own regulations for managed care as well as its own reimbursement practices that change the cost-cutting incentives of Medicaid managed care. The small cost increases that are observed probably come from spillovers from private insurance. The finding that the total health care share rose because of the managed care backlash is mostly driven by the behavior of the private health share.

Alternative Specification

For most of the estimates of the persistence parameter δ that I obtain in Table 1.3, δ is close to unity. Moreover, for some of these estimates, I cannot reject the null hypothesis that δ is

[13] In Figure 1.8, the counterfactual path of the health care share without backlash regulations first rises slightly during the recession of 2001, and then falls gradually below its original level by 2005. The fact that the difference between the counterfactual path and the observed path is slightly increasing over time is because backlash regulations affect the change of the health share of GSP, and therefore have a trend effect on the level of the health share of GSP. My estimates suggest that absent the backlash regulations, the health share of GDP would have been on a slight negative trend, and with the backlash regulations it was on a positive trend instead.

[14] In results not reported, I estimate equation 1.1 with log health share, log health expenditures per capita and log total health expenditures as dependent variables. The results are qualitatively similar, although the effects when the dependent variable is log expenditure or log expenditure per capita are of lower magnitude. Such a result is consistent with managed care allowing for health care costs to grow at roughly the same rate as the economy but not below this rate, either because patients valued the additional care sufficiently highly, or because customers were not particularly sensitive to price rises at or below the rate of economic growth.

equal to unity, and for the baseline specification the upper bound of the confidence interval for δ is unity. If δ is taken to be unity, equation (1.1) implies an equation of the form

$$\Delta P_{s,t} = \alpha_s + \lambda_t + \beta R_{s,t-1} + \gamma R_{s,t-1} \times HMO_s^{1995} + X_{s,t}'\eta + \varepsilon_{s,t} \tag{1.2}$$

Unlike equation (1.1), equation (1.2) is readily estimable by OLS and is more efficient when δ is actually equal to unity. It has an intuitive interpretation: it is just the regression of the change in the dependent variable on state characteristics, national trends, and the independent variables of interest. I now present my baseline results using the difference specification in Table 1.4, and present the rest of my results computed with the difference specification in Appendix I.

Table 1.4 shows the baseline results for the difference specification in column 2. The interaction coefficients and the forecasts are close to those produced using the dynamic panel specification. The forecast total health share of GDP is 11.9%, somewhat lower than for the dynamic panel specification, and it is significantly different from the observed 2005 level at 5% because the confidence interval of the prediction is much narrower. Such a forecast is consistent with the managed care backlash increasing the U.S. health share of GDP by 1.6 percentage points, which constitutes slightly more than the entire increase in the U.S. health share since 1993. In Table 1.4, I also present estimates of equation (1.1) obtained using the Arellano-Bond methodology. We see that the coefficient on the lagged dependent variable (δ) becomes smaller (0.86) and that the interaction coefficient is much smaller than in the nonlinear least squares specification (0.035) as per the prediction of Hausman, Hall and Kuersteiner (2007).

1.5.2 Robustness Checks

Elementary Robustness to Excluding Data Points

An elementary robustness check is to verify that my estimates are not sensitive to excluding individual states or groups of states from my sample. I therefore re-estimate equation (1.1) 50 times, dropping a different state each time, and look at the highest and lowest values attained by the interaction coefficient. I also repeat this exercise again 8 times, each time dropping a different region of the U.S. (New England, Mid-Atlantic, Southeast, Great Lakes, Plains, Southwest, Rocky Mountains, Pacific). The lower bound on the interaction coefficient is 0.09, and the upper bound is 0.13.[15]

Robustness to State Trends and Panel Covariates

Table 1.5 reestimates equation (1.1) when additional trends or control variables are added to the regression. Column 1 reestimates the baseline. Column 2 adds state-specific trends, a demanding robustness check (it effectively involves quadratic trends in the health share of GSP because of the persistence of the dependent variable). The interaction coefficient remains significant though it shrinks relative to the baseline (0.063), but the main effect of regulations becomes positive and the counterfactual forecast is about 9.9% of GDP, which is very low (and significantly different from the observed 2005 level even with the large prediction errors of the dynamic panel specification). Column 3 adds demographic covariates (log fractions of the population that are over 65, black, and female) to the baseline regression; the interaction coefficient shrinks slightly to 0.077 but

[15] In Appendix I, I carry out the same robustness check for the difference specification (1.2), with the addition that I also drop years from the sample, which I cannot do when estimating the dynamic panel model. I also plot the partial relation of the dependent variable against the HMO-regulations interaction.

remains significant at 1%. Column 4 adds log GSP. The interaction coefficient actually increases slightly and we do not see any substantial effects.[16] Finally, column 5 decomposes the raw count of regulations passed into individual counts for the 4 broad categories of regulations (access, appeals, mandates and provider regulations). This specification nests the baseline specification (which would obtain if the coefficients on all categories of regulations were the same), but allows different categories of regulations to affect the health care share differently. We see that the largest and statistically significant coefficients are on provider regulations, which suggests that regulations affecting the relationship between managed care and physicians (such as bans on financial incentives for physicians to treat less intensively, or any-willing-provider laws) were particularly important, followed by mandates for services that managed care especially tried to curtail (e.g. minimum maternity stays), while regulations expanding patients' access to physicians and procedures may have actually lowered the health care share. [17] The forecast is slightly lower than the baseline forecast.

Robustness to the Dynamic Structure of Regulations

An essential robustness check to ensure that my results are not being driven by mean reversion, or by various forms of reverse causation is to include leads and lags of my right-hand-side variables into the regression. Glied (2003) presents several theories of the rise in health care costs in the late 1990s and early 2000s, all of which argue that the health care cost slowdown in the 1990s was a product of a coincidence of transient factors (a low point in the underwriting cycle and strategic behavior of managed care firms during the health insurance market's transition to managed care in order to gain market share) that dissipated as the processes generating them reverted to the mean. Including leads and lags (together with contemporaneous effects) of the regulation variables into my regression helps control for mean reversion, and allows me to test an implication of the hypothesis that regulations are causing health care spending increases. Moreover, including leads and lags allows me to control for endogenous timing of the backlash regulations. For instance, if backlash regulations were passed in states with abnormally low health care share increases (because of aggressive cost containment that generated discontent), but then health care shares resumed rising (because of mean reversion), there would be a spurious positive correlation between lagged backlash regulations and current health care shares, and a spurious negative correlation between future backlash regulations and current health care shares. If the managed care backlash is causing changes in the health care share of GDP, it must be the case that when leads and lags of the regulations are included, the leads of the regulations are not significant conditional on the lags, while the lags are significant conditional on the leads. Table 1.6 presents the results for the dynamic panel specification (1.1). We see that the coefficients on the leads are an order of magnitude lower than the coefficients on the lags (the largest is 0.013). If only one lead, one lag and the contemporaneous effect are included, the interaction coefficient on the first lag is statistically significant. If two leads

[16] A more demanding robustness check would be to include not only the main effect of log GSP but also its interaction with HMO penetration in 1995. Then, the coefficient γ on the HMO-regulations interaction shrinks in magnitude and loses statistical significance. However, the resulting estimates greatly depend on outliers, and change drastically if a particular state is excluded. In Appendix I, I estimate this robustness check via median regression, which is more resistant to outliers than least squares is, and find results substantially closer to the baseline.

[17] However, all of these inferences should be interpreted with caution because the type of backlash regulations passed may be correlated with other aspects of the managed care backlash, such as adverse media coverage of managed care, which may have led it to curtail its cost containment practices.

I consider other ways of parametrizing the backlash regulations variable, e.g. counting the number of years for each state in which the number of backlash regulations increased (hence, the number of "backlash bills"). My results remain unchanged.

and lags with contemporaneous effects are included, each of the lag interaction coefficients is about 0.07, but neither is significant individually. Since multicollinearity becomes severe as leads and lags are added, I perform joint F-tests that all leads are zero and joint F-tests that all lags are zero. We see that the coefficients on leads are always jointly insignificant, while the coefficients on lags are jointly significant at 5% with only one lead and lag, and at 10% with two leads and lags. Therefore, we have some reassurance that it is the lags and not the leads that are driving my results.

Robustness to Other Health Insurance Regulations

A significant concern is that the managed care backlash in general, and backlash regulations in particular, proxy for other changes in the policy environment that cause the health spending share to rise. As discussed in Section 1.3, during the backlash period, other health insurance reforms that did not directly target HMO cost containment mechanisms – mandated benefits, small group and individual market insurance reforms, and Medicaid expansion – were being passed. It would be troubling both for my identification strategy and for my use of backlash regulations as a proxy for the intensity of the managed care backlash if controlling for these political changes in the health insurance environment significantly altered my baseline estimates, and it would be reassuring for my approach if accounting for other health insurance reforms did not appreciably change my results.

Table 1.7 attempts to address this concern by including these regulations in my baseline dynamic panel specification (1.1) alongside with the backlash regulations. Column 1 reproduces the baseline. Columns 2 through 5 add mandated benefits, small group reforms, individual market reforms and simulated Medicaid eligibility (both as levels and in interaction with HMO penetration) to the baseline regression, one at a time, respectively. Finally, Column 6 contains all the additional health insurance controls simultaneously (coefficients not reported). We see that the interaction coefficient on backlash regulations remains significant and unchanged in magnitude from the baseline specification, while the coefficients on the other health insurance reforms are insignificantly different from zero (with the exception of the individual market reforms). Moreover, the counterfactual forecasts under the hypothesis that no regulations were passed are similar to the baseline.

Robustness to Regional Disaggregation

To test my identification strategy further, I run my regressions using sub-state variation. While backlash regulations vary at the state by year level, I can use disaggregated data on HMO penetration, health spending and economic outcomes to add rich locational controls. Since only hospital spending data is available at the sub-state level (from the AHA Annual Survey), I can only look at total hospital spending, rather than at total health care spending. Moreover, because gross product data is not calculated for most sub-state units (in particular, for counties), I use county personal income as a measure of economic activity, which is different from gross output. Table 2.4 presents results when the unit of analysis is states, urban and rural counties of states agglomerated together (which I call MSU's), MSA's (with rural counties of a state combined into a single unit) and county zones (which are MSAs for urban counties, and agglomerations of 5 neighboring counties for rural counties). [18] We see that since personal income is smaller than gross output, the shares are

[18] Since each set of estimate requires performing nonlinear least squares with an increasing number of fixed effects, I cannot present results at the county level, which would require nearly 3000 fixed effects. I present results at the county level computed via OLS using specification 1.2 in Appendix I.

There are 97 MSUs because Alaska is classified as entirely rural, while New Jersey and Rhode Island are classified as entirely urban.

larger: the observed share of total spending out of U.S. personal income is 16.2%, and the forecast share without regulations is 14.9%. The first column reproduces the equivalent of my baseline specification with the new variables: the dependent variable is the change in health spending as a share of state personal income. The magnitude of the interaction coefficient is similar to the baseline estimate in Table 1.3, and the forecast share is practically and statistically significantly smaller than the observed share. Subsequent specifications use the change in hospital spending as a share of personal income as the dependent variable. Each specification at each unit of analysis includes unit fixed effects (thus, the MSA specification has MSA fixed effects), and since a lagged dependent variable is included, these fixed effects approximate linear unit trends, which is a very flexible way of controlling for many time-varying covariates (demographics, economic conditions) at a local level. The interaction coefficient is typically between 0.02 and 0.03, which is reasonable given that hospital spending is approximately a third of total personal care spending, with the exception of results for MSAs, where the interaction coefficient is 0.011. This is because looking at MSAs (with rural counties agglomerated into a single unit for analysis) downweights rural areas of the United States and increases the weight on urban areas. When we look at county zones (which do not downweight rural areas), the interaction coefficient is of the same magnitude as for states and MSUs. The counterfactual forecasts of hospital spending as a share of personal income are about 9% lower than the observed 2005 level for states and MSUs, but they are much closer to the observed 2005 levels for MSAs and county zones.

1.6 Results: Utilization and Health

It is interesting to examine what aspects of the health care production function did the managed care backlash affect to raise health care costs. Health care utilization and salaries are difficult to measure in the private sector because of a lack of centralized, consistent panel data. A partial solution is to look at hospital accounts and inventories, which have been systematically recorded for long periods of time by the American Hospital Association's Annual Survey. While these data describe only the health care production function of hospitals, we have seen in Table 1.8 that hospital spending constitutes over one-third of total personal health care spending, and that hospital spending has also risen with backlash regulations as total health care spending has. The AHA survey provides hospital-level data on aggregate measures of volume, such as admissions, inpatient days, payrolls and employment, allowing the computation of variables such as average hospital salaries and average lengths of stay per admission. In my analysis, I aggregate all variables to the state level.

Table 1.9 presents results of estimating the dynamic panel specification (1.1) for a variety of dependent variables measuring hospital expenditures, utilization, salaries, and technology adoption. Column 1 sets hospital expenditures as a share of personal income as the dependent variable. We see that the interaction coefficient is about 0.033 (which is reasonable given the fact that hospital expenditure is only 5.5% of personal income) and significant at 5%. Hospital expenditures as a fraction of personal income rose by over 12% (relative to the counterfactual level) during the managed care backlash. Column 2 shows that hospital payrolls as a fraction of personal income rose even more during the backlash than hospital expenditures did. It is interesting to attempt to understand how this rise in hospital payrolls as a share of personal income was allocated between

The county zones were constructed by ranking rural counties by their distance to the approximate centroid of the United States (-90 longitude, 38 latitude) and combining adjacent counties in that ranking within each state into groups of at most 5 counties in each group.

hospital employment (as a share of population) and hospital salaries (as a share of average personal income). Column 3 shows results for the association between backlash regulations and hospital employment as a share of population. We see that the interaction coefficient is very small and insignificant, and the counterfactual prediction is close to the baseline. Therefore, the increase in hospital payrolls as a share of GSP did not appear to have manifested itself in terms of higher hospital employment. Column 4 shows estimates for the association between backlash regulations and average hospital salaries as a fraction of state average personal income (measured in percentage points). We see that the interaction coefficient is significant at 5%, and that the observed 2005 average hospital salary as a fraction of the national average income is 10.5% higher than the counterfactual salary (although the confidence interval of this forecast is, as usual, extremely wide). Hence, there is suggestive evidence that most of the rise in payrolls as a share of income went into higher relative salaries for hospital workers rather than into increasing the fraction of the population in the hospital sector. This tentative finding is consistent with the estimates of Cutler, McCllelan and Newhouse (2000), which suggest that managed care reduced the salaries of medical providers.[19] Another potential source of cost increases during the managed care backlash (a source of cost savings during the managed care revolution) is length of stay (Glied 2000). Column 6 presents estimates for the association between length of stay and backlash regulations. The interaction coefficient in the length of stay specification is small and insignificant, but the main effect is positive and marginally significant, leading to a counterfactual 7.8% rise in the length of stay per admission. This is sizeable, though lower than the counterfactual hospital cost increase observed in column 1, suggesting that length of stay increases may have played a role in the cost increase observed during the managed care backlash. [20]

Given that we have found that there is a strong association between managed care regulation and health care cost growth, it is interesting to examine whether there were any discernible improvements in important measures of health that were associated with the backlash regulations. The literature on the impact of managed care on health outcomes and health care quality (summarized in Miller and Luft 1997 and Glied 2000) has not found substantial deteriorations or improvements in health arising from managed care. Theoretically, health could even improve with the introduction of managed care if some costly medical procedures were unnecessary or mildly harmful. In this paper, we confirm this basic result: the counterfactual path of health outcomes without managed care regulations is generally insignificantly different from the actually observed path, although the statistical uncertainty is high.

The health outcome I consider is all-population mortality. Unlike health outcomes such as

[19] It may be interesting to ask what happened to the earnings of physicians. While consistent panel data on physician incomes is difficult to obtain, I have computed average salaries in the ambulatory health sector (including physician offices, offices for outpatient treatment and home health care) using employment and compensation data from the Bureau of Economic Analysis. These salaries should be good proxies for physician income because they are the salaries of people employed by physicians, and whose productivity is directly related to that of physicians. The effects of the managed care backlash on ambulatory health sector salaries are similar but slightly stronger than those on hospital salaries (approximately a 14% counterfactual growth rate).

[20] In addition to the above analysis, I consider the effects of the managed care backlash on technological progress. The AHA also collects data on the different types of facilities that each hospital owns, which can be used to measure technological advancement (Cutler and Sheiner 1998, Baker and Phibbs 2002, Finkelstein 2007, Acemoglu and Finkelstein 2008). I measure technological advancement by the weighted number of facilities per million people in a state, with each facility receiving a weight proportional to the fraction of hospitals lacking such a facility in 1995, a weighting scheme pioneered by Baker and Spetz (1999). Dynamic panel specification estimates are very noisy (and the usual sign pattern of the coefficients on the regulation variables is reversed). However the counterfactual forecast implies that the managed care backlash resulted the number of facilities per million being 15.8% higher than without the backlash. Difference specification estimates of this relationship (presented in Appendix I, Table 1.16) result in a similar counterfactual forecast and in a positive, though insignificant, interaction coefficient.

the incidence or severity of major diseases (which may increase as the population ages because of improved life expectancy), mortality for a fixed age group is an unambiguous indicator of a poor health outcome. However, mortality in the privately insured population under 65 is (fortunately) low, which makes it difficult to estimate any potential effects of backlash regulations precisely. Column 7 of Table 1.9 presents evidence on the effects of managed care regulation on overall all-cause age-adjusted mortality rate for the entire U.S. population (per 100,000) using the dynamic panel specification (1.1). We see that the forecast mortality rate in 2005 if backlash regulations were absent is slightly lower than the observed rate, although the confidence intervals are very wide. The interaction coefficient is negative, suggesting that backlash regulations may have lowered mortality in states with high HMO penetration relative to states with low HMO penetration, but the main effect is large, leading to mortality actually being less in the counterfactual scenario without the managed care backlash than was actually observed in 2005. The confidence intervals on the forecasts are too wide to calculate any meaningful upper bounds on the degree to which mortality may have fallen with backlash regulations.

1.7 Political Determinants of Regulations and Instrumental Variables Estimation

Since the managed care backlash was partially mediated by the political system through the passage of the backlash regulations, it is natural to ask whether political variables explain backlash regulations. Moreover, if it can be argued that these political variables could have impacted the health care market only through the passage of backlash regulations, it would be possible to test my identification strategy further by using the political variables as instruments. Obtaining valid instrumental variables estimates for the effect of the managed care backlash on health care costs that matched with the OLS estimates presented in the baseline results would be reassuring confirmation of the validity of the central findings of my paper.

The political variables I will be using for most of my analysis will be numbers of years of Democratic control of the state governorships, upper houses of state legislatures, and lower houses of state legislatures since 1994 (since the first large wave of backlash regulations came in 1995). There is good reason on the basis of the health policy literature to believe that Democrats were more favorably disposed to backlash regulations than Republicans were. When the U.S. House of Representatives voted on the Bipartisan Consensus Managed Care Improvement Act (H.R. 2723) in 1999 (also known as the Norwood-Dingell Act), which would have imposed a federal version of the backlash regulations (including managed care liability for poor health outcomes resulting from denials of care), all but five Democrats voted for passage, while nearly three-fourths of Republicans voted against passage (Poole and Rosenthal 2012). Brodie et al. (1998) and Gray et al. (2007) provide evidence that self-identified Democrats were more likely to support backlash regulations. However, there were exceptions: the Texas Health Care Liability Act, one of the most comprehensive pieces of backlash legislation, was passed in Texas in 1998 with the strong support of the Republican governor, George W. Bush. A more technical reason for looking at the Democratic control variables is that other political variables that may conceivably affect backlash regulations – such as interest group activity of supporters or opponents of these regulations – are much more poorly measured because of data problems and because of lack of clarity in what constitutes interest group activity in principle. On the other hand, whether the Democrats or the Republicans are in control of a particular branch of state government is very easy to ascertain.

Since there are three parts of the state government whose control I can assign to a party, I create multiple variables for Democratic control of the various combinations of parts of the state government. Moreover, motivated by the example of Texas, I include interactions of the Democratic control variables with an indicator that the state in question is a Southern state, since the relative Democratic propensity to support backlash regulations there is very different than in the rest of the country. I describe in detail my procedure for parametrizing the Democratic control variables in Appendix III, and I provide tentative evidence that Democratic control increases the passage of backlash regulations. The individual coefficients are imprecisely estimated and do not present a particular pattern, but it is possible to see that any configuration of Democratic control increases the propensity to pass backlash regulations outside the South.

Can the association between the backlash regulations and the Democratic control variables be used to implement an instrumental variables strategy for estimating the effect of the managed care backlash on health care costs? It can under the condition that the Democratic control variables affect health care costs exclusively through the backlash regulations. In principle, Democratic control of branches of the state government could affect health care costs through the passage of other health insurance regulations, or through other legislation that affects the economy as a whole, and through it, the health care share of GSP. From the analysis in Section 1.5, we see that other health insurance regulations do not affect the relationship between backlash regulations and health care costs and tend not to significantly affect health care costs once backlash regulations are accounted for. The concern that non-health-related legislation favored by Democrats affects the health care share remains, but is less plausible because such indirect effects would have to be very large to have meaningful impacts.

A way to further refine this identification strategy would be to find a variable that would affect the ability of Democrats to pass backlash regulations and use only the interaction of that variable with the Democratic control variables as an instrument (while accounting for the main effects of the Democratic control variables in the structural regression). I argue that such a variable is physician dominance of health interest groups, which is used in Gray et al. (2007), and which is measured as the fraction of health lobby registrations by primary care clinic organizations. From the discussion in Section 1.2, we see that physicians were vocal opponents of managed care cost containment practices, both because these practices interfered with the clinical practices that they were accustomed to and that were parts of their training, and because managed care adversely impacted medical provider salaries (Cutler, McClellan and Newhouse 2000, Section 1.6 this paper). Gray et al. (2007) finds that physician dominance in the early period of the backlash is correlated with the subsequent passage of backlash regulations in a cross section of states. Therefore, we should expect physician dominance of health interest groups to make it easier for state governments to pass backlash regulations, all else the same. On the other hand, it is unlikely that physician dominance of health interest groups could affect the health care share except through the passage of health care regulations. Physician dominance could be endogenous to the health share, but using measures of physician dominance for the pre-backlash period and the early backlash period should ameliorate this problem.

Hence, a more conservative identification strategy would be to instrument backlash regulations with interactions between Democratic control variables and the pre-period measure of physician dominance of health interest groups, while including the Democratic control variables without physician dominance interactions as controls in the first and second stage. The identification assumption becomes that the only way in which Democratic control of the state government could differentially affect the health care share as a function of pre-period physician dominance of health interest groups is through the passage of backlash regulations. Since it is very implausible that Democrats would have a propensity to pass legislation that does not affect the health care market directly in a way that varies with physician dominance of interest groups, we no longer have the concern that Democrats may have passed non-health-related legislation with indirect effects on the health care market. It still could be the case that physician groups differentially influenced Democrats' ability to pass other health insurance regulations that were not related to the backlash. However, we have seen in Table 1.7 that of the major health insurance regulations passed during the backlash period, only the backlash regulations appear to affect health care costs. Finally, including the Democratic control main effects in my regression ensures that it is the interaction relationship between the Democratic controls and pre-period physician dominance that is providing the variation to instrument backlash regulations, rather than the Democratic-physician dominance interactions acting as a proxy for Democratic controls. In Appendix III, I provide evidence that variation in these interaction coefficients is sufficient to help explain the passage of backlash regulations, even controlling for the main effects of Democratic controls. Moreover, I provide tentative evidence that the Democratic-physician dominance interactions increase the passage of backlash regulations. Once again, the full specification contains many coefficients whose signs do not reveal a pattern. Simple parametrizations of the backlash regulations-Democratic control-physician dominance relationship exhibit the intuitive signs for the estimates, but these estimates tend not to be statistically significant.

Table 1.10 presents instrumental variables estimates of the dynamic panel specification 1.1 based on the two instrumental variables strategies I propose.[21] To estimate Table 1.10, I use nonlinear GMM by exploiting the exclusion restrictions implied by the excluded political instruments and the regressors assumed to be exogenous. Column 1 reproduces the OLS results. The subsequent columns contain instrumental variables results for various combinations of instruments. Since there are many instruments in each regression, I do not present the first stages, but simply write what groups of instruments are included. In addition to the second stage results, I present the results of tests for underidentification, overidentification and exogeneity of the instrumented regressors. First, to show that my instruments are relevant to the instrumented regressors, I present p-values of tests of excluded instruments for the regression of regulations on all variables presumed to be exogenous. Second, since I have more instruments than instrumented regressors in every specification, I present the p-value of the Hansen overidentification test on the objective function, which tests the null hypothesis that the model is overidentified (all instruments are exogenous conditional on one instrument being exogenous). Finally, I present the Hausman test for the endogeneity of the instrumented variables, conditional on my instruments being exogenous.

Column 2 instruments backlash regulations and backlash regulations interacted with 1995 HMO penetration using the Democratic control variables only, both as main effects and interacted with the South dummy. We see that the interaction coefficient is virtually the same magnitude and significance as in the baseline specification, and the counterfactual forecast is slightly larger (12.5%) and significantly different from the observed 2005 level. This is because the standard error on the lagged dependent variable coefficient is substantially smaller than in the baseline specification, although the standard errors on the other regressors are similar to the baseline. We also see that both first stages are significant, that the Hansen test cannot reject overidentification of the model (so the instruments are not mutually contradictory), and the Hausman test cannot reject the null hypothesis that the backlash regulations are exogenous (which is not surprising, given the similarity of the coefficient estimates). Column 3 executes the more conservative identification strategy and instruments the two backlash regulation variables with Democratic control-physician dominance interactions only (with or without the South dummy). The main effects of the Democratic control variables are included as exogenous variables in the regression in order to isolate the variation coming from the interaction terms. We see that now the interaction coefficient is slightly larger than the baseline (0.116), and the counterfactual forecast is 12.4%. We also see that the

[21] I present similar estimates for the difference specification 1.2 via 2SLS in Appendix I.

excluded instruments (the Democrat-physician dominance interactions) are significant in the first stage even after controlling for the Democratic main effects. As in Column 2, the Hansen test fails to reject overidentification, and the Hausman test fails to reject exogeneity of the regulations.

Therefore, under either of my identification assumptions, we see that endogeneity in the backlash regulations is not likely to be a problem for my analysis, and in particular, that the backlash regulations are most likely exogenous with respect to shocks to the health care share of GSP. We also obtain a tentative story for an aspect of the political system's role in the passage of backlash regulations: the Democratic party, at least outside the South, was relatively more likely to pass such regulations than the Republican party was, and the presence of physician-dominated health interest groups increased this party differential in backlash regulation passage. [22]

1.8 Conclusion

This paper finds that the managed care backlash of the late 1990s, as measured by state regulation of managed care cost containment practices, has increased the U.S. health care share of GDP by over 1 percentage point, and accounts for much of the growth in the health care share of GDP since the health care cost growth stagnation of the 1990s. This result is robust to a variety of specification checks, which, in particular, rule out alternative explanations based on neglected geographic heterogeneity, mean reversion, and confounding with other health insurance policies. There is suggestive evidence that the backlash operated mostly by raising the costs of privately-insured patients through higher provider salaries and more intensive utilization (longer lengths of stay), with a possible increase in the rate of technological change. I further show that there were no statistically significant mortality improvements caused by the managed care backlash. Finally, I present evidence that political variables can explain part of the variation in the backlash regulations, and exploit this observation to execute an instrumental variables strategy.

Given that the magnitude of the cost increase that I attribute to the managed care backlash is comparable to the sizes of the major U.S. public health insurance programs, it is worth studying the phenomenon of the backlash in greater detail. While a great deal of qualitative research has been done on the backlash in the health policy literature, to my knowledge, this is the first paper that investigates the backlash in public economics. It is important to understand precisely what components of the managed care backlash (media or regulatory) had the largest effects on health care costs, and what channels did the backlash operate through to raise the U.S. health care share. It is also important to understand why some states experienced a much stronger level of backlash than did others. Additionally, my finding highlights the importance of studying health care cost control in the private sector, especially given the Affordable Care Act's emphasis on using private insurance to achieve universal coverage.

Furthermore, my findings emphasize the importance of studying the virtues and defects of different managed care cost containment mechanisms. We cannot quantify the many inconveniences – reduced choice of treatment strategy, inability to see a doctor one has been accustomed to, unpredictability of utilization review committees – that managed care created for its patients, and therefore cannot trade them off against the cost savings. Suggestive evidence from looking at the

[22] I do not make any normative claims on whether passing the backlash regulations was welfare-improving or welfare-deteriorating. While I find evidence that backlash regulations increased health care costs, I cannot find statistically significant evidence on the benefits of backlash regulations for health and peace of mind of patients, and therefore, cannot judge whether the benefits exceeded the costs. In fact, my mortality estimates alone are compatible with substantial health benefits from the backlash regulations.

effects of different types of regulations hints that some of these hardships could have been regulated away without substantial cost, and that most of the cost savings from managed care occurred from its ability to influence physicians rather than patients. In light of the inclusion of managed care into the Affordable Care Act through ACOs, it is imperative to understand what particular aspects of the managed care program created value for its customers so that it could be possible to improve on the managed care model in the future.

1.9 Figures

Figure 1.1 (1.1)

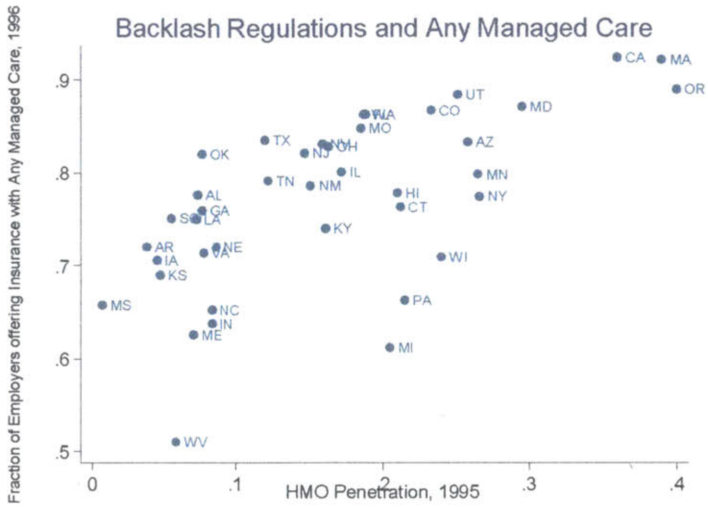


Figure 1.2 (1.2)

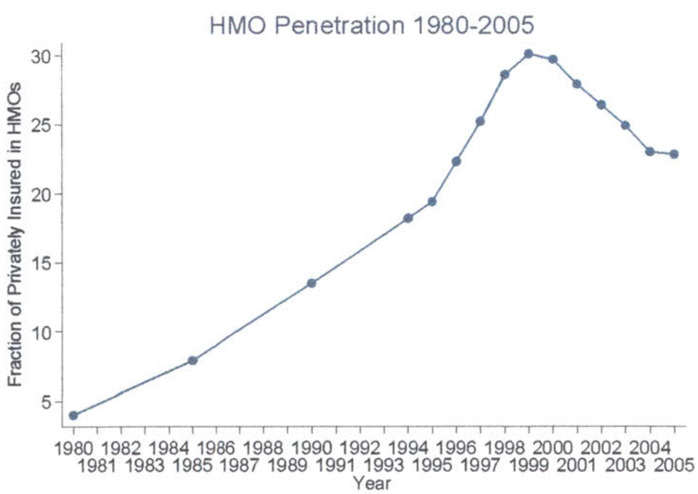


Figure 1.3 (1.3)

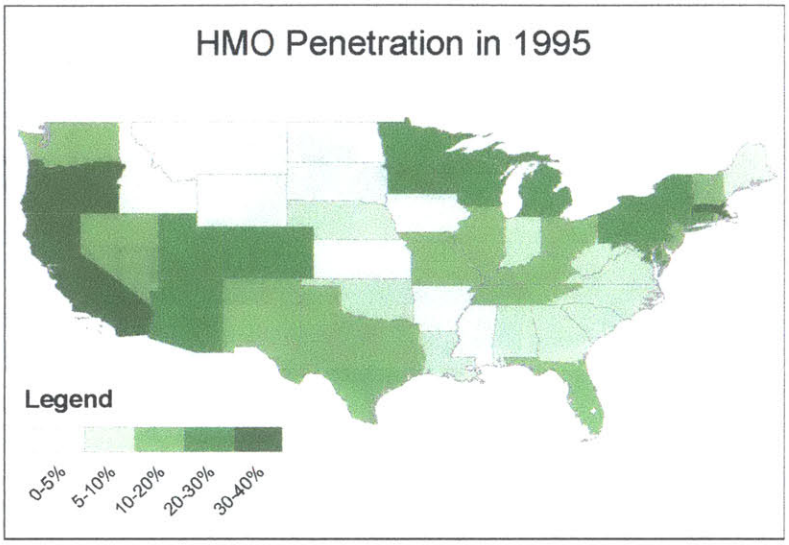


Figure 1.4 (1.4)

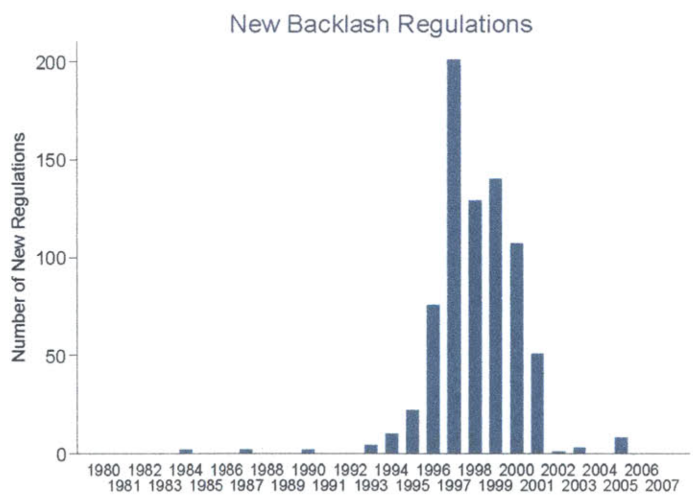


Figure 1.5 (1.5)

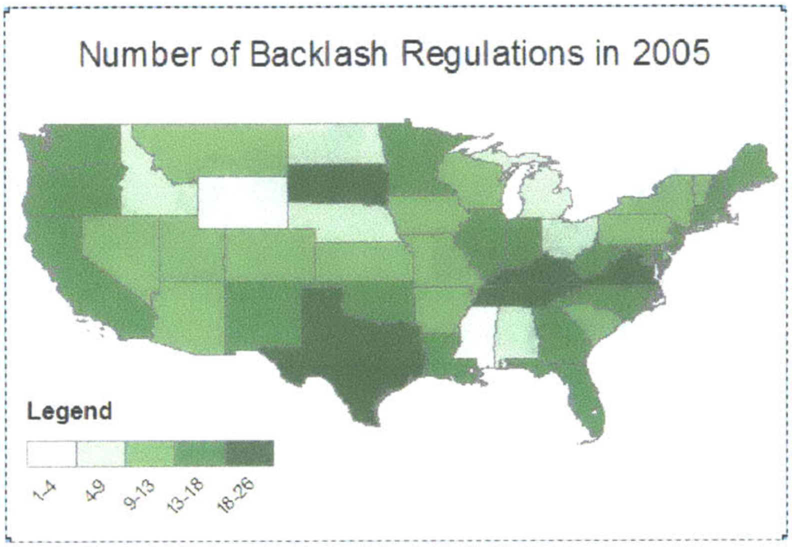


Figure 1.6 (1.6)

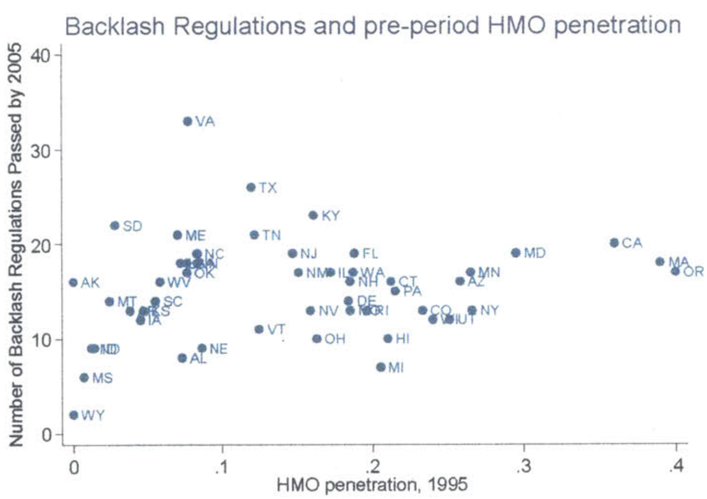


Figure 1.7 (1.7)

Distribution of Counterfactual Forecast of Health Care Share, %

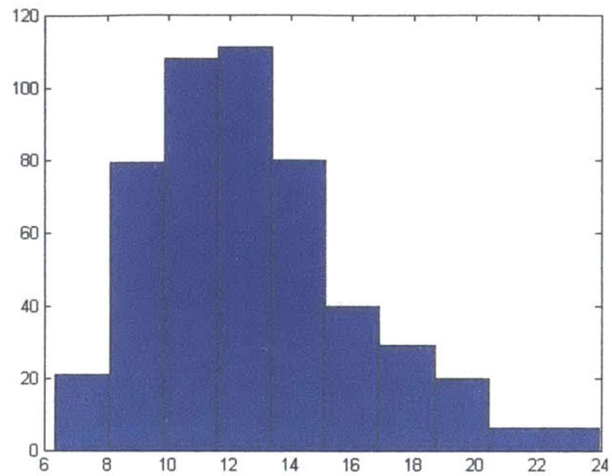
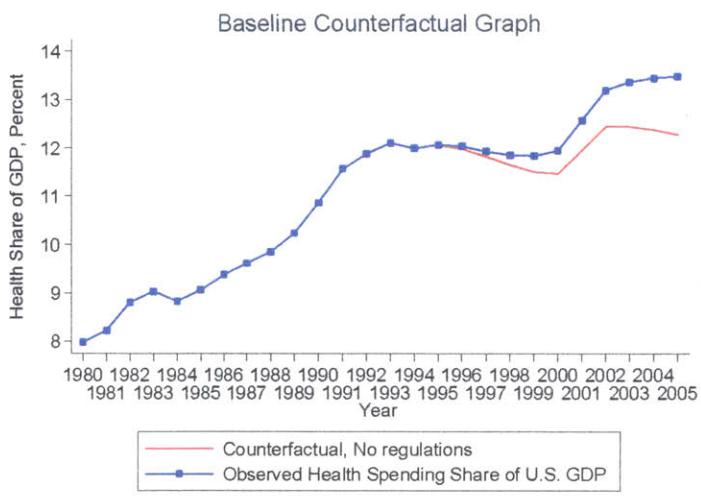


Figure 1.8 (1.8)



1.10 Tables

Table 1.1

Backlash Regulation Type (Fine Grouping)	Coarse Grouping	Number of Regulations of given Type
Comp. Consumer Rights	Access	68
Continuity of Care	Access	40
Direct Access, OB/GYN	Access	48
Direct Access, other	Access	21
Emergency Care Coverage	Access	39
Emergency Room	Access	3
Emergency Prudent Lay Person	Access	23
Ombudsman	Access	21
Specialist as PCP	Access	10
Standing Ref. To Specialist	Access	28
Insurer Liability	Appeals	14
Independent External Review of Denials	Appeals	58
Liability, Financial: Enrollee	Appeals	16
Liability: Provider Contracts	Appeals	26
Point of Service	Appeals	21
Diabetes Supplies	Mandates	54
Hospital Stay after Childbirth	Mandates	42
Inpatient Care after Mastectomy	Mandates	22
Post-Mastectomy Breast Reconstruction	Mandates	10
Off-label Prescription Drug Use	Mandates	18
Any Willing Provider	Provider	16
Ban All Products Clauses	Provider	6
Ban on Financial Incentives	Provider	38
Ban on Gag Clauses	Provider	57
Freedom of Choice	Provider	9
Medical Director Requirements	Provider	26
Report Cards	Provider	27

(1.1)

Table 1.2

Summary Statistics, State-Level Data

VARIABLES	(1) N	(2) Mean	(3) SD
Personal Health Share of GSP	50	14.28	2.408
Change, Personal Health Share of GSP	50	0.104	0.357
Backlash Regulations	50	15.22	5.300
HMO Penetration in 1995, %	50	14.54	10.17
PAHAshare	50	4.942	1.125
Hospital Payroll Exp. as Share of Personal Income	50	2.617	0.613
Hospital Employment as Share of Population	50	1.739	0.334
Average Hospital Salary as Share of Income per Capita	50	150.7	20.59
Average Length of Stay	50	6.746	1.276
Number of Facilities per Million, Rare-Weighted	50	629.3	347.6
Backlash Regulations, Access	50	8.360	3.445
Backlash Regulations, Appeals	50	1.120	1.003
Backlash Regulations, Mandates	50	2.820	1.480
Backlash Regulations, Provider	50	2.920	1.469
Small Group Full Reforms	50	0.720	0.454
Indiv. Mrkt. Full Reform	50	0.180	0.388
Mandated Benefits	50	15.40	5.440
Log Gross State Product	50	11.90	1.046
All-Cause Mortality Rate per 100,000	50	818.3	86.80

(1.2)

Data on state regulation of managed care obtained from the National Conference of State Legislatures. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract, originally from Interstudy. Data on health expenditures and GSP from CMS. Data on hospital expenditures, payrolls, employment, lengths of stay and facilities from the AHA Annual Survey. Data on other health insurance regulations from "State Legislative Health Care and Insurance Issues" by BCBS. Data on mortality from the CDC.

Table 1.3

Baseline Estimates *Dynamic Panel Specification*				
	(1)	(2)	(3)	(4)
	Total Share	Private Share	Medicare Share	Medicaid Share
Lag. DV	.929**	.897**	.932**	.931**
	(.035)	(.034)	(.063)	(.060)
Regs (T-1)	-.007	-.009+	-.001	.003
	(.006)	(.004)	(.001)	(.002)
Regs (T-1) X HMO (1995)	.093**	.085**	.010	-.011
	(.025)	(.017)	(.006)	(.008)
Observed level in U.S. (2005)	13.48	8.59	2.59	2.29
Forecast w/o Regulations	12.28	7.82	2.49	2.16
95% CI Upper Bound	20.40	12.32	6.35	5.06
95% CI Lower Bound	7.85	5.17	1.27	.96
Number of Obs.	550	550	550	550
Number of Clusters	50	50	50	50
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

(1.3)

Each column presents results from estimating equation (1.1) with suitable covariates. Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract, originally from Interstudy. Data on health expenditures and GSP from CMS. Private health expenditures are defined as the difference between total expenditures and Medicare and Medicaid expenditures.To compute the 2005 forecast, I draw 500 independent observations from the distribution of the coefficient vector and dynamically simulate the counterfactuals of all regulation variables being set to zero.

Table 1.4

Alternative Specifications			
	(1)	(2)	(3)
	Total Share	Diff Total Share	Total Share
Lag. DV	.929**		.858**
	(.035)		(.034)
Regs (T-1)	-.007	-.008	-.000
	(.006)	(.006)	(.009)
Regs (T-1) X HMO (1995)	.093**	.112**	.048
	(.025)	(.023)	(.037)
Estimation Method	NLS	OLS	AB
No. Observations	550	550	550
No. Clusters	50	50	50
R^2		.38	
Observed U.S. level (2005)	13.48	13.48	13.48
Forecast w/o Regulations	12.28	11.92*	12.68
95% CI Upper Bound	20.40	13.13	18.32
95% CI Lower Bound	7.85	10.53	9.06
State FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

(1.4)

Column 1 presents results from estimating equation (1.1) via the Hausman-Pinkovskiy procedure with suitable covariates, column 2 presents results from estimating equation (1.2) via OLS, and column 3 presents results from estimating equation (1.1) via Arellano-Bond. Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract, originally from Interstudy. Data on health expenditures and GSP from CMS. To compute the 2005 forecast, I draw 500 independent observations from the distribution of the coefficient vector and dynamically simulate the counterfactuals of all regulation variables being set to zero.

Table 1.5

Robustness Checks *Dynamic Panel Specification* *Dep. Var. is Total Health Share of GSP*					
	(1)	(2)	(3)	(4)	(5)
Spec.	Base Line	State Trends	Demo Graph.	GDP	4 Groups Counts
Lag. DV	.929**	1.133**	.964**	1.047**	.946**
	(.035)	(.007)	(.040)	(.038)	(.040)
Regs (T-1)	-.007	.005**	-.005	-.010+	
	(.006)	(.001)	(.005)	(.005)	
Regs (T-1) X HMO (1995)	.093**	.063**	.077**	.098**	
	(.025)	(.003)	(.025)	(.023)	
Access Regs (T-1) X HMO (1995)					-.114+
					(.069)
Appeals Regs (T-1) X HMO (1995)					.417
					(.367)
Mandates Regs (T-1) X HMO (1995)					.357+
					(.195)
Provider Regs (T-1) X HMO (1995)					.435**
					(.141)
Observed level in U.S. (2005)	13.48	13.48	13.48	13.48	13.48
Forecast w/o Regulations	12.28	9.90*	12.17	11.79	12.19
95% CI Upper Bound	20.40	13.30	23.23	28.92	25.10
95% CI Lower Bound	7.85	6.84	6.65	3.82	6.95
Number of Obs.	550	550	550	550	550
Number of Clusters	50	50	50	50	50
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

(1.5)

Each column presents results from estimating equation (1.1) with suitable covariates. Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on health expenditures and GSP are from CMS. Column 3 includes demographic controls for (log) proportion of the population over 65 (in Medicare), proportion black and female, proportion black and male, proportion white and male, and proportion white and female. Column 4 includes log GSP per capita as a control, and column 5 breaks down regulations into 4 groups as in 1.1 (main effects are suppressed). To compute the 2005 forecast, I draw 500 independent observations from the distribution of the coefficient vector and dynamically simulate the counterfactuals of all regulation variables being set to zero.

Table 1.6

Leads and Lags *Dynamic Panel Specification* *Dep. Var. is Total Health Share of GSP*				
	(1)	(2)	(3)	(4)
Lag Structure:	-1	0	-1/1	-2/2
Lag. DV	.929**	.922**	.926**	.926**
	(.035)	(.034)	(.036)	(.036)
Regs (T-2) X HMO (1995)				.069
				(.067)
Regs (T-1) X HMO (1995)	.093**		.176*	.065
	(.025)		(.075)	(.104)
Regs X HMO (1995)		.092**	-.100	-.051
		(.024)	(.102)	(.111)
Regs(T+1) X HMO (1995)			.010	-.004
			(.081)	(.142)
Regs(T+2) X HMO (1995)				.013
				(.097)
P-value Leads are Zero			.85	.98
P-value Lags are Zero			.03	.09
Observed level in U.S. (2005)	13.48	13.48	13.48	13.48
Forecast w/o Regulations	12.28	12.40	12.52	12.73
95% CI Upper Bound	20.40	19.93	21.41	21.33
95% CI Lower Bound	7.85	8.10	7.70	7.82
Number of Obs.	550	550	550	550
Number of Clusters	50	50	50	50
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

(1.6)

Each column presents results from estimating equation (1.1) with suitable covariates and state and year fixed effects. Standard errors clustered by state are in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on health expenditures and GSP are from CMS. All regressions contain main effects that are suppressed. To compute the 2005 forecast, I draw 500 independent observations from the distribution of the coefficient vector and dynamically simulate the counterfactuals of all regulation variables being set to zero.

Table 1.7

Robustness to Other Health Insurance Regulations *Dynamic Panel Specification* *Dep. Var. is Total Health Share of GSP*						
	(1)	(2)	(3)	(4)	(5)	(6)
Other Reg:		Other Mandated Benefits	Small Group Reform	Indiv. Mrkt. Reform	Mcd Simltd Elig.	All Other Regs.
Lag. DV	.929**	.937**	.925**	.937**	.937**	.952**
	(.035)	(.033)	(.036)	(.034)	(.036)	(.032)
Regs (T-1)	-.007	-.018	-.009	-.005	-.009	-.016
	(.006)	(.013)	(.005)	(.006)	(.006)	(.013)
Regs (T-1) X HMO (1995)	.093**	.153**	.098**	.088**	.095**	.148**
	(.025)	(.046)	(.024)	(.025)	(.025)	(.046)
Oth. Reg. (T-1)		.031	.119	-.558**	.659	
		(.027)	(.104)	(.157)	(.913)	
Oth. Reg. (T-1) X HMO (1995)		-.142	-.251	2.392*	-.085	
		(.116)	(.720)	(1.160)	(5.466)	
Observed level in U.S. (2005)	13.48	13.48	13.48	13.48	13.48	13.48
Forecast w/o Regs	12.28	12.18	12.32	12.13	12.36	11.94
95% CI Upper Bound	20.40	20.32	20.35	19.92	20.78	20.18
95% CI Lower Bound	7.85	7.52	7.90	7.76	7.74	7.17
Number of Obs.	550	550	550	550	550	550
Number of Clusters	50	50	50	50	50	50
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

(1.7)

Each column presents results from estimating equation (1.1) with suitable covariates and state and year fixed effects. Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on health expenditures and GSP are from CMS. Data on mandated benefits, small group reforms and individual market reforms is obtained from Blue Cross Blue Shield's "State Legislative Health Care and Insurance Issues." The mandated benefits variable is the sum of mandated benefits. Following Simon (2000) I consider a state to have passed a small group reform if it has guaranteed issue, guaranteed renewal and rating reform, and the individual market reform is coded similarly. Data on simulated Medicaid eligibility from Kosali Simon. To compute the 2005 forecast, I draw 500 independent observations from the distribution of the coefficient vector and dynamically simulate the counterfactuals of all regulation variables being set to zero.

Table 1.8

Robustness to Regional Disaggregation *Dynamic Panel Specification* *Dep. Var. is Total Hospital Share of Unit Personal Income and Total Health Share of State Personal Income in Column 1*					
	(1)	(2)	(3)	(4)	(5)
Lag. DV	.974**	.856**	.869**	.859**	.847**
	(.030)	(.052)	(.060)	(.044)	(.059)
Regs (T-1)	-.006	-.000	-.002	-.000	-.003
	(.006)	(.003)	(.002)	(.002)	(.002)
Regs X HMO (LB) (T-1)	.089**	.026+	.035*	.011	.024*
	(.020)	(.014)	(.014)	(.008)	(.010)
No. Observations	550	550	1067	4971	9143
No. Units	50	50	97	455	835
No. Clusters	50	50	50	50	50
Observed U.S. level (2005)	16.2	5.44	5.44	5.44	5.44
Forecast w/o Regulations	14.87	5.04	5.06	5.22	5.31
95% CI Upper Bound	17.43	5.95	6.21	6.16	6.32
95% CI Lower Bound	12.36	4.22	4.08	4.39	4.42
Unit of Analysis	State	State	MSU	MSA	Zone

(1.8)

Each column presents results from estimating equation (1.1) with suitable covariates and unit and year fixed effects. Standard errors in parentheses, clustered at the state level. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of county population enrolled in HMOs in 1995 (aggregated up when necessary) is obtained from an original dataset compiled by Laurence Baker, originally from Interstudy. Data on hospital total expenditures is obtained from the AHA Annual Survey, and data on county personal income (aggregated up when necessary) is obtained from the BEA. In column (1), the dependent variable is total health spending as a share of state personal income.

Table 1.9

	(1)	(2)	(3)	(4)	(5)	(6)
Other Outcomes *Dynamic Panel Specification*						
Dep. Var.	Hospital Expend. Share of GSP	Payroll Expenditures Share of GSP	Hospital Emplmt. Share of Pop.	Hospital Avg. Sal. Share of GSP p/c	Hospital Inp. Days Share of Pop.	Adj. Mort. Rate
Lag. DV	.877**	.875**	.883**	.851**	.521**	.868**
	(.049)	(.053)	(.045)	(.063)	(.147)	(.028)
Regs (T-1)	-.001	-.001	-.000	-.012	.007+	.124
	(.002)	(.001)	(.000)	(.071)	(.004)	(.127)
Regs (T-1) X HMO (1995)	.033*	.020**	.004	.740*	.016	-.255
	(.014)	(.007)	(.003)	(.346)	(.025)	(.290)
Observed level in U.S. (2005)	4.56	2.39	1.62	147.69	6.48	808.40
Forecast w/o Regulations	4.06	2.12	1.56	133.73	6.01	795.51
95% CI Upper Bound	7.15	4.02	2.70	285.06	14.42	1133.41
95% CI Lower Bound	2.54	1.24	.99	75.84	3.66	563.87
Forecast Growth, %	12.22	12.83	3.67	10.43	7.79	1.62
Number of Obs.	550	550	550	550	550	550
Number of Clusters	50	50	50	50	50	50
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

(1.9)

See Table 2.3. Data on all dependent variables is obtained from the AHA Annual Survey. Data on health expenditures and GSP obtained from CMS.

Table 1.10

Instrumental Variable Estimates			
Dynamic Panel Specification			
Dep. Var. is Total Health Share of GSP			
	(1)	(2)	(3)
	OLS	GMM	GMM
Lag. DV	.929**	.929**	.962**
	(.035)	(.001)	(.001)
Regs (T-1)	-.007	-.008**	-.013*
	(.006)	(.002)	(.006)
Regs (T-1) X HMO (1995)	.093**	.095**	.116**
	(.025)	(.012)	(.025)
Excl. P-val., Regs		.00	.00
Hansen P-val.		.93	.92
Hausman P-val. vs. Baseline		.65	.34
P-val. Dem. Exog. Vars.			.00
Observed level in U.S. (2005)	13.48	13.48	13.48
Forecast w/o Regulations	12.28	12.45**	12.41**
	(3.13)	(.27)	(.35)
Dem. Insts.		Yes	Yes
Phys. Dom X Dem Inst.		No	No
Dem. Cntrls. in Stage 2		No	No
No. Observations	550	550	550
No. Clusters	50	50	50
State FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

(1.10)

Each column presents results from estimating equation (1.1) via nonlinear GMM using the exclusion restrictions implied by the instruments. Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on health expenditures and GSP from CMS. Data on Democratic control obtained from the Statistical Abstract of the United States. Data on physician dominance of health interest groups (fraction of health lobby organizations by primary care clinic organizations obtained as personal communication from Virginia Gray. Column 3 contains the Democratic controls (with and without interaction with South dummy) included as exogenous variables but not as instruments. The coefficient estimates for these variables are not reported for brevity.

1.11 Appendix I: Difference Specification Results

In this appendix, I describe the results that I obtain when I use the difference specification (1.2). Using this specification is tantamount to imposing that the coefficient on the lagged dependent variable in specification (1.1) is equal to unity. Such an assumption substantially decreases the standard errors of my forecasts, allowing me to conclude that many of these forecasts are statistically significantly different from the observed levels of the variables in question in 2005. Setting the lagged dependent variable coefficient to unity also enables estimation via OLS. However, the difference specification imposes an additional restriction on the data relative to my main specification (1.1), and this restriction is rejected for some of the specifications that I estimate, so I present results for it in an appendix.

Figure 1.9 shows an elementary set of robustness checks in which I plot the partial relation between the growth of the private health share of GSP and the interaction of backlash regulations with HMO penetration in 1995, as well as bounds from estimations of equation (1.2) dropping individual states, regions, or years. The bounds are reasonably tight around the baseline estimate of 0.112 (the lower bounds are not less than 0.099, while the upper bounds are not greater than 0.134), suggesting that the results are not driven by outlier states, or outlier years, or by any single region (which is important to verify, since HMO penetration differs greatly by region).

Table 1.11 shows analogous results to Table 1.3 (Baseline) for the difference specification. The interaction coefficients and the forecasts are close to those produced using the dynamic panel specification. The forecast total and private health shares of GDP are lower than for the dynamic panel specification, and they are significantly different from the observed 2005 levels at 5% because the standard errors of the prediction are much lower. The forecast Medicaid health spending share is substantially lower than in Table 1.3, but is not significantly different from the observed 2005 level even with the reduced standard errors.

Table 1.12 presents the same results as Table 1.5 (Robustness Checks) for the difference specification (1.2). The results are very similar to the dynamic panel results, except that the counterfactual when state trends are added is higher (12.71%) and the interaction coefficient is no longer significant, though of the same magnitude as in the dynamic panel specification. In addition, I perform a more demanding robustness check in Column 5 by including the interaction between log GSP and HMO penetration. The coefficient on the regulations-HMO interaction becomes insignificant and shrinks to about 0.4. Upon examination of this result, I find that the estimates strongly depend on outlier observations; excluding the state of Oregon would raise the interaction coefficient to 0.77. Therefore, column 6 reestimates the specification in column 5 using median regression, which is more robust to outliers. The interaction coefficient remains insignificant, but rises in magnitude 50% to 0.6, and the counterfactual estimate becomes 12.71%, the same as when state trends are included, and slightly lower than when GSP is included as a main effect only.

Table 1.13 presents analogous results to Table 1.6 (Leads and Lags) for the difference specification (1.2). The results are less clear, with the magnitude of the lead coefficients equal to one-half or two-thirds that of the lag coefficients, and with no individual coefficient being significant. The joint F-tests suggest that the lags are significant while the leads are not in the two-lag specification (in which the lead coefficients are the largest in magnitude), but that both groups of coefficients are jointly insignificant in the one-lag specification. It is likely that the mean reversion present in Table 1.13 can be explained by the slight negative correlation between the level of the health care share and its change over time, which is explicitly captured by the dynamic panel specification in Table 1.6.

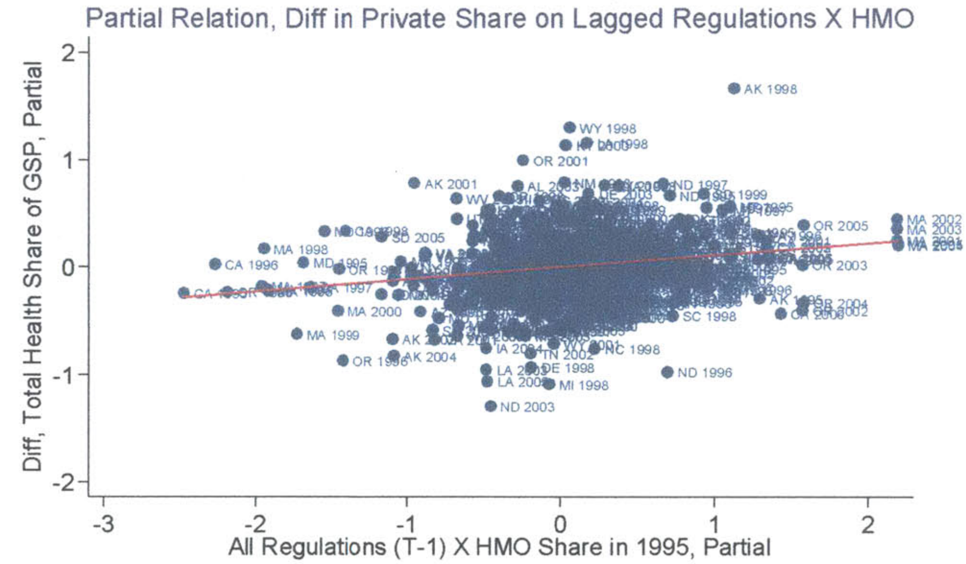
Table 1.14 replicates the analysis in Table 1.7 (Other Health Insurance Regulations) for the difference specification (1.2). We see that virtually nothing changes, except the counterfactual forecast of the health share of GDP rises to 12.6% when all the other health insurance regulations considered are controlled for.

Table 1.15 repeats and enhances the analysis in Table 1.8 (Regional Disaggregation). In addition to the regressions from Table 1.8, it also includes a county-level regression, which involves too many fixed effects to be computed via nonlinear optimization. The results of Table 1.15 are somewhat stronger than in Table 1.8, with larger and more significant interaction coefficients, and lower counterfactual forecasts. We see that, just as in Table 1.8, the results for MSAs are weaker than all the other results, but the results when the unit of analysis is the county are as strong as the results for states and MSUs. Therefore, it is the fact that looking at MSAs skews the sample towards cities, rather than the fact that MSAs are too disaggregated, that explains the low impacts of backlash regulations for MSAs. However, these results should be treated with caution because from Table 1.8, it appears that the lagged dependent variable coefficient is less than unity, so the results from Table 1.8 should be preferable.

Table 1.16 replicates and enhances the results in Table 1.9 using the difference specification (1.2) (Other Outcomes). The interaction coefficient estimates and the are broadly similar, although the implied counterfactual growth rates are much larger. The counterfactual increase in the hospital share is now 26%, and te counterfactual increase in the hospital payroll share is now 20%. All these counterfactual forecasts are significantly different from zero. The counterfactual increase in average hospital salaries (as a fraction of average income) is now 19%, and also significantly different from zero. I also estimate the association between backlash regulations and the number of facilities in hospitals per capita in a state (a proxy for technological intensity). The interaction coefficient in the technology regression is now positive (as would be predicted if the backlash had its effects on technology through a change in the behavior of managed care) and the counterfactual forecast is that the number of facilities per capita is 17.92% higher than without the backlash. Both the main effect and the interaction coefficient in the mortality regression are also positive, suggesting that backlash regulations increased mortality, but once again, the standard errors of the forecast are too large for meaningful value-of-life calculations.

Table 1.17 repeats the second-stage instrumental variables analysis in Table 1.10 for the difference specification using conventional two-stage least squares rather than nonlinear GMM. The results change very little. The standard errors on the coefficients increase, which leads to the Democratic main effects in Column 3 no longer being significant.

Figure 1.9 (1.9)



Jackknife Bounds from Dropping States: [.106 , .125].
Jackknife Bounds from Dropping Regions: [.099 , .134].
Jackknife Bounds from Dropping Years: [.104 , .124]

Table 1.11

Baseline Estimates *Difference Specification*				
	(1)	(2)	(3)	(4)
	Diff Total Share	Diff Private Share	Diff Medicare Share	Diff Medicaid Share
Regs (T-1)	-.008	-.011*	-.002	.004+
	(.006)	(.004)	(.001)	(.002)
Regs (T-1) X HMO (1995)	.112**	.106**	.013*	-.007
	(.023)	(.015)	(.005)	(.009)
R^2	.38	.28	.59	.23
Observed level in U.S. (2005)	13.48	8.59	2.59	2.29
Forecast w/o Regulations	11.92*	7.42*	2.59	1.92
95% CI Upper Bound	13.13	8.31	2.89	2.37
95% CI Lower Bound	10.53	6.38	2.24	1.39
Number of Obs.	550	550	550	550
Number of Clusters	50	50	50	50
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

(1.11)

Each column presents results from estimating equation (1.2) with suitable covariates. Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract, originally from Interstudy. Data on health expenditures and GSP from CMS. Private health expenditures are defined as the difference between total expenditures and Medicare and Medicaid expenditures.To compute the 2005 forecast, I draw 500 independent observations from the distribution of the coefficient vector and dynamically simulate the counterfactuals of all regulation variables being set to zero.

Table 1.12

Robustness Checks *Difference Specification* *Dep. Var. is Diff., Total Health Share of GSP*							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Spec.	Base Line	State Trends	Demo Graph.	GDP	GDP X HMO	GDP X HMO Median Reg.	4 Regs Counts
Regs (T-1)	-.008	-.006	-.005	-.010+	-.005	-.005	
	(.006)	(.013)	(.004)	(.005)	(.008)	(.009)	
Regs (T-1) X HMO (1995)	.112**	.068	.083**	.077**	.041	.060	
	(.023)	(.075)	(.024)	(.026)	(.045)	(.057)	
Log GSP (T-1)				2.228**	1.989**	1.609*	
				(.419)	(.455)	(.745)	
Log GSP X HMO (T-1)					1.520	.957	
					(1.658)	(2.220)	
Access Regs (T-1) X HMO (1995)							-.007
							(.089)
Appeals Regs (T-1) X HMO (1995)							.187
							(.379)
Mandates Regs (T-1) X HMO (1995)							.186
							(.219)
Provider Regs (T-1) X HMO (1995)							.437*
							(.174)
R^2	.38	.43	.40	.42	.42	.	.38
Observed level in U.S. (2005)	13.48	13.48	13.48	13.48	13.48	13.48	13.48
Forecast w/o Regulations	11.92*	12.71	12.20*	12.97	13.23	12.71	12.03*
95% CI Upper Bound	13.13	14.93	13.27	14.29	14.68	14.62	13.45
95% CI Lower Bound	10.53	10.57	11.04	11.48	11.80	10.93	10.69
Number of Obs.	550	550	550	550	550	550	550
Number of Clusters	50	50	50	50	50	50	50
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

(1.12)

Each column presents results from estimating equation (1.2) with suitable covariates. Standard errors clustered by state in parentheses. See notes to Table 1.5. Column 5 adds an interaction between log GSP and HMO penetration, and column 6 reestimaetes column 5 using median regression. Column 7 breaks down regulations into 4 groups as in 1.1 (main effects are suppressed). To compute the 2005 forecast, I draw 500 independent observations from the distribution of the coefficient vector and dynamically simulate the counterfactuals of all regulation variables being set to zero.

Table 1.13

Leads and Lags *Difference Specification* *Dep. Var. is Diff., Total Health Share of GSP*				
	(1)	(2)	(3)	(4)
Lag Structure:	-2	0	-1/1	-2/2
Regs (T-2) X HMO (1995)				.092
				(.065)
Regs (T-1) X HMO (1995)	.112**		.122	.025
	(.023)		(.086)	(.122)
Regs X HMO (1995)		.106**	.036	.044
		(.018)	(.131)	(.125)
Regs(T+1) X HMO (1995)			-.062	-.011
			(.086)	(.139)
Regs(T+2) X HMO (1995)				-.066
				(.104)
R^2	.38	.37	.38	.38
P-value Leads are Zero			.75	.79
P-value Lags are Zero			.22	.02
Observed level in U.S. (2005)	13.48	13.48	13.48	13.48
Forecast w/o Regulations	11.92*	12.26*	12.15	11.81
95% CI Upper Bound	13.13	13.42	14.30	13.77
95% CI Lower Bound	10.53	10.90	10.30	9.57
Number of Obs.	550	550	550	550
Number of Clusters	50	50	50	50
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

(1.13)

Each column presents results from estimating equation (1.2) with suitable covariates and state and year fixed effects. Standard errors clustered by state are in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on health expenditures and GSP are from CMS. All regressions contain main effectsthat are suppressed. To compute the 2005 forecast, I draw 500 independent observations from the distribution of the coefficient vector and dynamically simulate the counterfactuals of all regulation variables being set to zero.

Table 1.14

Robustness to Other Health Insurance Regulations *Difference Specification* *Dep. Var. is Diff., Total Health Share of GSP*						
	(1)	(2)	(3)	(4)	(5)	(6)
Other Reg:		Other Mandated Benefits	Small Group Reform	Indiv. Mrkt. Reform	Mcd Simltd Elig.	All Other Regs.
Regs (T-1)	-.008	-.019*	-.010	-.008	-.008	-.020*
	(.006)	(.009)	(.006)	(.006)	(.006)	(.010)
Regs (T-1) X HMO (1995)	.112**	.149**	.113**	.112**	.106**	.146**
	(.023)	(.042)	(.022)	(.024)	(.024)	(.047)
Oth. Reg. (T-1)		.037+	-.002	-.040	.014	
		(.020)	(.081)	(.292)	(.679)	
Oth. Reg. (T-1) X HMO (1995)		-.101	1.000	-.565	2.254	
		(.098)	(.615)	(2.075)	(3.578)	
R^2	.38	.38	.38	.38	.38	.39
Observed level in U.S. (2005)	13.48	13.48	13.48	13.48	13.48	13.48
Forecast w/o Regs	11.92*	12.40	12.05*	11.85*	12.06*	12.59
95% CI Upper Bound	13.13	13.87	13.30	13.05	13.27	14.13
95% CI Lower Bound	10.53	10.80	10.58	10.47	10.67	10.92
Number of Obs.	550	550	550	550	550	550
Number of Clusters	50	50	50	50	50	50
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

(1.14)

Each column presents results from estimating equation (1.2) with suitable covariates and state and year fixed effects. Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on health expenditures and GSP are from CMS. Data on mandated benefits, small group reforms and individual market reforms is obtained from Blue Cross Blue Shield's "State Legislative Health Care and Insurance Issues." The mandated benefits variable is the sum of mandated benefits. Following Simon (2000) I consider a state to have passed a small group reform if it has guaranteed issue, guaranteed renewal and rating reform, and the individual market reform is coded similarly. Data on simulated Medicaid eligibility from Kosali Simon. To compute the 2005 forecast, I draw 500 independent observations from the distribution of the coefficient vector and dynamically simulate the counterfactuals of all regulation variables being set to zero.

Table 1.15

Robustness to Regional Disaggregation *Difference Specification* *Dep. Var. is Diff., Hospital Expenditure Share of Unit Personal Income and Diff., Total Health Share of State Personal Income in Column 1*						
	(1)	(2)	(3)	(4)	(5)	(6)
Regs (T-1)	-.009*	.000	-.002	-.002	-.002	-.001
	(.004)	(.003)	(.002)	(.002)	(.002)	(.001)
Regs (T-1) X HMO (LB) (1995)	.099**	.036**	.049**	.023*	.035**	.033**
	(.015)	(.011)	(.010)	(.010)	(.010)	(.005)
No. Observations	550	550	1067	4971	9143	24592
No. Units	50	50	97	455	835	2469
No. Clusters	50	50	50	50	50	50
R^2	.54	.35	.2	.03	.02	.06
Observed U.S. level (2005)	16.2	5.44	5.44	5.44	5.44	5.44
Forecast w/o Regulations	14.91*	4.58*	4.54*	5.18	4.91*	5.07*
95% CI Upper Bound	15.84	5.17	5.09	5.75	5.38	5.37
95% CI Lower Bound	13.83	3.89	3.92	4.50	4.36	4.72
Unit of Analysis	State	State	MSU	MSA	Zone	County

(1.15)

Each column presents results from estimating equation (1.2) with suitable covariates and unit and year fixed effects. Standard errors in parentheses, clustered at the state level. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of county population enrolled in HMOs in 1995 (aggregated up when necessary) is obtained from an original dataset compiled by Laurence Baker, originally from Interstudy. Data on hospital total expenditures is obtained from the AHA Annual Survey, and data on county personal income (aggregated up when necessary) is obtained from the BEA. In column (1), the dependent variable is total health spending as a share of state personal income.

Table 1.16

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Other Outcomes *Difference Specification*							
	Diff Hospital Expend. Share of GSP	Diff Payroll Expend. Share of GSP	Diff Hospital Emplmt. Share of Pop.	Diff Hospital Avg. Sal. Share of GSP p/c	Diff Hospital Inp. Days Share of Pop.	Diff Number of Facilities Per Million	Diff Adj. Mort. Rate
Regs (T-1)	-.000	-.001	-.001	.000	.000	.016	.017
	(.003)	(.001)	(.001)	(.080)	(.003)	(.825)	(.140)
Regs (T-1) X HMO (1995)	.042**	.026**	.004	1.031**	.028	3.314	.249
	(.012)	(.007)	(.004)	(.332)	(.023)	(2.788)	(.467)
R^2	.22	.58	.10	.58	.09	.11	.41
Observed level in U.S. (2005)	4.56	2.39	1.62	147.69	6.48	469.19	808.40
Forecast w/o Regulations	3.61*	1.98*	1.64	124.10*	5.80	397.87	801.74
95% CI Upper Bound	4.20	2.33	1.82	138.95	6.75	530.48	829.10
95% CI Lower Bound	2.94	1.57	1.43	107.37	4.79	240.26	769.95
Forecast Growth, %	26.02	20.82	-1.40	19.01	11.66	17.92	.83
Number of Obs.	550	550	550	550	550	550	550
Number of Clusters	50	50	50	50	50	50	50
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

(1.16)

See Table 1.11. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on all dependent variables is obtained from the AHA Annual Survey. Data on health expenditures and GSP obtained from CMS. Technology count is weighted count of all facilities in state, with weights following Baker and Spetz (1999) equal to percentage of hospitals lacking the facility in 1995.

Table 1.17

Instrumental Variable Estimates *Difference Specification* *Dep. Var. is Diff., Total Health Share of GSP*			
	(1)	(2)	(3)
	OLS	IV	IV
Regs (T-1)	-.008	-.000	-.022
	(.006)	(.008)	(.015)
Regs (T-1) X HMO (1995)	.112**	.094+	.151*
	(.022)	(.052)	(.060)
Excl. P-val., Regs		.00	2.02e-16
Excl. P-val., Regs X HMO		3.51e-28	2.49e-12
P-val. Dem. Exog. Vars.			.66
Hansen P-val.		.55	.66
Hausman P-val. vs. Baseline		.58	.68
Observed level in U.S. (2005)	13.48	13.48	13.48
Forecast w/o Regulations	11.92*	11.45	12.73
95% CI Upper Bound	13.07	13.85	15.84
95% CI Lower Bound	10.59	8.91	9.18
Dem. Insts.		Yes	No
Phys. Dom X Dem Inst.		No	Yes
Dem. Cntrls. in Stage 2		No	Yes
No. Observations	550	550	550
No. Clusters	50	50	50
State FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

(1.17)

Each column presents results from estimating equation (1.2) via two-stage least squares.Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on health expenditures and GSP from CMS. Data on Democratic control obtained from the Statistical Abstract of the United States. Data on physician dominance of health interest groups (fraction of health lobby organizations by primary care clinic organizations obtained as personal communication from Virginia Gray. Column 3 contains the Democratic controls (with and without interaction with South dummy) included as exogenous variables but not as instruments. The coefficient estimates for these variables are not reported for brevity.

1.12 Appendix II: More on the Dynamic Panel Specification

This section draws heavily on Hausman and Pinkovskiy (2013).

We are interested in estimating the coefficients δ and γ in the equation

$$P_{s,t} = \delta P_{s,t-1} + X_{s,t}\gamma + \varepsilon_{s,t} \tag{1.A1}$$

under the assumption

$$E\left(\varepsilon_{s,t}|X_{\sigma,\tau}\right) = 0, \text{ for all } s,t,\sigma \text{ and } \tau, \text{ and } E\left(\varepsilon_{s,t}\varepsilon_{\sigma,\tau}\right) = 0 \text{ for all } (s,t) \neq (\sigma,\tau) \tag{1.AA1}$$

without using instrumental variables. (Some of the components of $X_{s,t}$ may be state or year fixed effects). I assume that we have covariates and dependent variables from period 1 to period T, and we also observe the dependent variable in period 0, so that we have exactly T observations.

Note that assumption (1.AA1) has two parts: first, all covariates are assumed to be strictly exogenous, and second, the error term is assumed to be serially uncorrelated. The uncorrelatedness assumption is necessary so that $E\left(\varepsilon_{s,t}|P_{s,0}\right) = 0$, since $P_{s,0} = f\left(\varepsilon_{s,-1}, ...\varepsilon_{s,-t}.....\right)$, a function of all prior error terms.

We can recursively substitute equation (1.A1) into itself to obtain the equation

$$\begin{aligned} P_{s,t} &= \delta P_{s,t-1} + X_{s,t}\gamma + \varepsilon_{s,t} \\ &= \delta^t P_{s,0} + \sum_{\tau=0}^{\tau=t-1} \delta^\tau L^\tau X'_{s,t}\gamma + \sum_{\tau=0}^{\tau=t-1} \delta^\tau \varepsilon_{s,t-\tau} \\ &= \delta^t P_{s,0} + \sum_{\tau=0}^{\tau=t-1} \delta^\tau L^\tau X'_{s,t}\gamma + \eta_{s.t} \end{aligned} \tag{1.A2}$$

Then, it is clear that $E\left(\eta_{s,t}|X_{\sigma,\tau}\right)\ \forall\sigma\forall\tau = E\left(\eta_{s,t}|P_{s,0}\right) = 0$ because of the assumptions on $\varepsilon_{s,t}$, so nonlinear least squares estimation of (1.A2) will yield consistent estimates of δ and γ. Specifically, $\left(\delta^{NLS}, \gamma^{NLS}\right)$ will solve

$$\left(\delta^{NLS}, \gamma^{NLS}\right) = \arg\min_{\delta,\gamma} \sum_{s,t} \left(P_{s,t} - \delta^t P_{s,0} - \sum_{\tau=0}^{\tau=t-1} \delta^\tau L^\tau X_{s,t}\gamma\right)^2$$

We can weaken assumption (1.AA1) considerably. First, we can dispense with the uncorrelatedness component by using lagged values of $X_{s,t}$ as instruments. In the case when the error term $\varepsilon_{s,t}$ is correlated, we have one endogeneous variable $(P_{s,0})$, and we have at least as many excluded instruments from the lagged values of $X_{s,t}$ as there are independent variables. (Specifically, if we have K regressors and T periods, we have $K(T-1)$ excluded instruments). Under the assumption

$$E\left(\varepsilon_{s,t}|X_{\sigma,\tau}\right) = 0, \text{ for all } s,t,\sigma \text{ and } \tau \tag{1.AA2}$$

and defining $\hat{X}_{s,t}$ as the vector $[X_{s,t}, X_{s,t-1}, X_{s,t-2}...X_{s,1}]$, we have the exclusion restriction

$$E\left(\hat{X}'_{s,t}\left(P_{s,t} - \delta_0^t P_{s,0} - \sum_{\tau=0}^{\tau=t-1} \delta_0^\tau L^\tau X'_{s,t}\gamma_0\right)\right) = \sum_{\tau=0}^{\tau=t-1} \delta_0^\tau E\left(\hat{X}'_{s,t}\varepsilon_{s,t-\tau}\right) = 0$$

satisfied. Note that we used strict exogeneity here.

Then, the GMM estimator of (δ_0, γ_0) solves

$$\left(\delta^{GMM}, \gamma^{GMM}\right) = \arg\min_{\delta,\gamma} \left(P - \delta_0^t P_0 - \sum_{\tau=0}^{\tau=t-1} \delta_0^\tau L^\tau X\gamma_0\right)' \hat{X}\Xi\hat{X}' \left(P - \delta_0^t P_0 - \sum_{\tau=0}^{\tau=t-1} \delta_0^\tau L^\tau X\gamma_0\right)$$

where Ξ is the optimal weighting matrix.

We can further weaken assumption (1.AA1) by assuming that $X_{s,t}$ is only predetermined. Specifically, we assume

$$E\left(\varepsilon_{s,t}|X_{\sigma,\tau}\right)=0 \text{ unless } \sigma=s \text{ and } \tau>t \tag{1.AA3}$$

It is clear that the previous exclusion restriction fails because $E\left(X_{s,t}'\varepsilon_{s,1}\right)$ cannot be assumed to be zero for any $t>1$, so $\left(\delta^{GMM},\gamma^{GMM}\right)$ are biased and inconsistent. We do however have the exclusion restriction $E\left(X_{s,1}\varepsilon_{s,\tau}\right)=0$ for all $\tau\geq 1$, which gives us K instruments. Since we have $K+1$ endogenous variables, this is not enough, but if we have covariates for period zero, then we can get an instrument vector $X^{-}=[X_{s,0},X_{s,1}]$. (We can similarly use any covariates $X_{s,t}$ for $t\leq 0$ if they are available). We then have at least $2K\geq K+1$ instruments, and we can proceed as before.

Reformulated with X^{-}, the following exclusion restriction holds:

$$E\left(X_{s,t}^{-\prime}\left(P_{s,t}-\delta_0^tP_{s,0}-\sum_{\tau=0}^{\tau=t-1}\delta_0^{\tau}L^{\tau}X_{s,t}'\gamma_0\right)\right)=\sum_{\tau=0}^{\tau=t-1}\delta_0^{\tau}E\left(X_{s,t}^{-\prime}\varepsilon_{s,t-\tau}\right)=0$$

and the new GMM estimator solves

$$\left(\delta^{GMM},\gamma^{GMM}\right)=\arg\min_{\delta,\gamma}\left(P-\delta_0^tP_0-\sum_{\tau=0}^{\tau=t-1}\delta_0^{\tau}L^{\tau}X\gamma_0\right)'\left(X^{-}\right)\Xi\left(X^{-}\right)'\left(P-\delta_0^tP_0-\sum_{\tau=0}^{\tau=t-1}\delta_0^{\tau}L^{\tau}X\gamma_0\right)$$

GMM estimation is computationally intensive, and occasionally fails to converge if rich covariates are included. Therefore, I perform Hausman tests to show that we cannot statistically distinguish estimators based on assumption (1.AA1) from those based on assumptions (1.AA2) and (1.AA3).

Table 1.18 presents several versions of the baseline specification. We see that failing to include lagged dependent variables or failing to difference the dependent variable (use the growth rate rather than the log) results in noisy estimates that suggest that backlash regulations lowered health care costs. OLS or Arellano-Bond estimation yields a positive and significant interaction coefficient, but the interaction coefficient is smaller than in the baseline specification, and the persistence coefficient on the lagged dependent variable is about 0.8-0.87, suggesting less than complete persistence. Column 4 presents estimation of equation (1.1) using the nonlinear least squares method of Hausman and Pinkovskiy (2013). We see that the persistence coefficient is 0.93, barely distinguishable from unity at 5%. Columns 5 and 6 estimate (1.1) by nonlinear GMM in order to check the robustness of the dynamic panel results to weaker assumptions about the exogeneity of the error term. Column 5 instruments for the initial value using lags of the independent variables (backlash regulations and their interaction with pre-period HMO penetration), and Column 6 instruments for all the independent variables using backlash regulations and their interactions from 1995 (the first year of the sample) and 1994. We see that the estimates in columns 5 and 6 are statistically indistinguishable from those in column 4, and we verify it formally by using the Hausman test.

Table 1.18

Specification Analysis							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total Share	Total Share	Total Share	Total Share	Total Share	Total Share	Diff Total Share
Lag. DV		.799***	.858***	.929***	.933***	.927***	
		(.032)	(.034)	(.035)	(.005)	(.001)	
Regs (T-1)	.002	-.006	-.000	-.007	-.015	-.008***	-.008
	(.026)	(.007)	(.009)	(.006)	(.010)	(.003)	(.006)
Regs (T-1) X HMO (1995)	-.196*	.050	.048	.093***	.124***	.073***	.112***
	(.101)	(.031)	(.037)	(.025)	(.035)	(.014)	(.023)
No. Observations	550	550	550	550	550	550	550
No. Clusters	50	50	50	50	50	50	50
R^2	.94						.38
P-value of Hausman Test against Col. 4					.76	1	
Observed 2005 Dep. Var. in U.S.	13.48	13.48	13.48	13.48	13.48	13.48	13.48
Forecast 2005 Dep. Var. in U.S. if no Regulations	14.27**	13.16	12.68	12.28	12.47	12.88**	11.92**
95% CI Upper Bound	14.46	17.10	18.32	20.40	14.90	13.18	13.13
95% CI Lower Bound	14.08	10.54	9.06	7.85	10.18	12.59	10.53
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

(1.18)

Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on health expenditures and GSP from CMS. Column 1 presents OLS estimates without the lagged dependent variable. Column 2 presents OLS estimates of equation (1.1). Column 3 presents Arellano-Bond estimates of equation (1.1) with all feasible instruments used. Column 4 presents the nonlinear least squares estimates of equation (1.1) under Assumption (1.AA1). Column 5 presents GMM estimates of equation (1.1) under the assumption (1.AA2). Column 6 presents GMM estimates of equation (1.1) under assumption (1.AA3). Column 7 presents OLS estimates of the difference specification equation (1.2).

1.13 Appendix III: More on Instrumental Variables

To parametrize the extent of Democratic control during the backlash period, I create 7 variables in total for the 7 combinations of Democratic control that can obtain in any given year.[23] Each variable is the number of years since 1994 that the state government experienced the particular configuration of Democratic control.The omitted variable is the number of years since 1994 that Democrats have controlled no part of the state government. Since the dependent variable is the total number of regulations outstanding in a given state by a given year, it makes sense to look at the cumulative number of years of Democratic control rather than at whether Democrats control the state government at the given point in time. The main motivation for such a parametrization is that if support for backlash regulations was partisan, then the Democratic control variables span the possible combinations of partisan control of the state government, and therefore, flexibly capture any influences of partisan control.

Column 1 of Table 1.19 shows the regression of backlash regulations on the 7 Democratic control variables. We see that while the coefficients of these variables have different signs, one year of Democratic control of any combination of the branches of a state government increases the number of backlash regulations.[24] However, none of the coefficients is significant, and the 7 Democratic coefficients are insignificant jointly. The explanation for this failure of statistical significance is that the relative support of the Democratic party for backlash regulations was not homogeneous across the United States. Motivated by the Texas example, in which a Republican governor supported backlash regulations, in column 2, I present the regression of backlash regulations on the 7 Democratic control variables as main effects, and on 7 interactions between Democratic control variables and a dummy variable indicating that the state in question is a Southern state. The specification in column 2 explicitly allows for differences in relative Democratic support for backlash regulations between the South and the rest of the U.S. [25] We see that an additional year of Democratic control of any configuration of state government branches increases the number of backlash regulations outside the South (with the exception of just the control of the lower house), but not necessarily in the South. Most importantly, we see that the 14 Democratic controls with interactions for the South are jointly significant, and therefore, help explain the passage of backlash regulations.

Tables 1.20 and 1.21 provide some intuition concerning the relationship between backlash regulations, Democratic control, and pre-period physician dominance. Since in a specification with Democratic main effects, Democrat-South interactions, Democrat-physician dominance interactions, and Democrat-physician dominance-South interactions, there are 28 different coefficients, I present these coefficients in columns 3 and 4 of Table 1.19 but do not discuss them. Instead, Table

---

[23] Hence, these variables are the numbers of years since 1994 that Democrats have controlled 1) the governorship, 2) the upper house, 3) the lower house, 4) both the upper house and the lower house, 5) both the governorship and the upper house, 6) both the governorship and the lower house, and 7) the governorship, the upper house and the lower house all together.

[24] To see this, consider a configuration of Democratic control, e.g. governor and upper house. An extra year of this configuration of Democratic control will have an impact on regulations equal to the coefficient for a Democratic governor, plus the coefficient for a Democratic upper house, plus the coefficient for the combination of a Democratic governor and a Democratic upper house. We see that this sum is greater than zero. A similar analysis can be done for all other configurations.

[25] There are two reasons why the relationship between Democratic control of the state government and the passage of backlash regulations could have been different in the South as compared to the rest of the United States. First, the Democratic and Republican parties were much more similar in the South than they were nationally in the 1990s – many Southern Republicans had earlier been Democrats, and many Southern Democrats were maintaining their party affiliation by force of habit rather than because of substantial agreement with the nationwide Democratic party. Second, the 1990s saw a transition from virtually solid Democratic state government in the South to a substantial presence of Republicans, which created further policy convergence because of political competition.

1.20 provides the p-values that all regressors are zero, and the p-values that all regressors with physician dominance interactions are zero for four specifications that explain backlash regulations with the variables discussed. We see that controlling for differences in Democratic relative support for backlash regulations between the South and the rest of the U.S. is crucial for joint significance of all regressors. We also see that the Democrat-physician dominance interactions (with and without South dummy interactions) are significant even when Democrat main effects are included in the regression. In fact, these interactions are significant at 10% even when South dummy interactions are not included (they are significant at 1% when they are included). Therefore, the political variables I have identified have explanatory power for backlash regulations, and specifically, there appear to be statistically significant differential effects on relative Democratic propensity to pass backlash regulations as a function of pre-period physician dominance, so there is variation to exploit for my second, more conservative identification strategy. Unfortunately, Table 1.20 does not provide good information for the direction of the effects: on whether Democrats are more inclined to support backlash regulations relative to Republicans, and on how pre-period physician dominance affects this relative support. Therefore, Table 1.21 presents the coefficients for regressions when each Democratic control indicator is analyzed separately. Each regression has four variables: the Democratic control in question, the Democratic control interacted with the South dummy, the Democratic control interacted with pre-period physician dominance, and the triple interaction of all three variables. No variable in any regression is statistically significant, so this exercise should be interpreted as, at most, illustrative. We see that in all the regressions, the Democratic main effect is positive, suggesting Democrats pass more backlash regulations outside the South than Republicans do, as expected. The Democrat-South interaction is negative in all but one of the specifications, suggesting this effect is decreased or reversed in the South, also as expected. The Democrat-physician dominance interaction is positive in all but one specification, suggesting that physician dominance of health interest groups increased the relative Democratic propensity to pass backlash regulations outside the South. This is expected, because it is likely that the efforts of physician groups and of Democrats to pass backlash regulations were supermodular (since physician groups could mobilize grassroots support for the regulations, while Democrats could vote the regulations into law). Hence, physician interest groups were more capable of getting backlash regulations passed when Democrats were in office than when Republicans were. Finally, the triple interaction coefficient is sometimes positive and sometimes negative. However, there is no reason to expect this coefficient to be of a particular sign; in the South, physician interest groups may have been especially helpful in increasing the differential Democratic propensity to pass backlash regulations because this differential propensity was low to begin with, or they may have had less of a differential effect on Democratic passage of regulations because both parties were sufficiently similar to begin with. Hence, we have evidence that there exists experimental variation in backlash regulations that I can exploit for an instrumental variables strategy, and we have suggestive evidence for a story that backlash regulations were passed more frequently by Democrats than by Republicans, with this differential increased in the presence of physician-dominated health interest groups.

Table 1.19

Determinants of Regulations *Dep. Var. is # Backlash Regulations*				
	(1)	(2)	(3)	(4)
Governor	.442	.301	.607+	.441
	(.336)	(.310)	(.360)	(.400)
Upper Hse	.331	.170	1.068	-.075
	(.381)	(.297)	(1.241)	(.552)
Lower Hse	.456	-.061	3.307+	-3.483*
	(.588)	(.367)	(1.694)	(1.542)
State. Leg.	-.499	.314	-3.985	3.917*
	(.736)	(.485)	(2.489)	(1.696)
Ctrl. All	-.139	-.963	2.077	-6.243**
	(1.029)	(.779)	(3.157)	(1.997)
Gov.+ UH	-.444	-.254	-1.498	-.201
	(.608)	(.481)	(1.377)	(.605)
Gov.+ LH	.148	.915	-1.626	5.928**
	(.863)	(.636)	(2.496)	(1.815)
Governor X South		.721		-1.039*
		(.480)		(.508)
Upper Hse X South		1.428		21.965**
		(1.720)		(3.751)
Lower Hse X South		3.317**		-41.584*
		(1.088)		(19.545)
State. Leg. X South		-5.203*		20.185
		(2.334)		(22.784)
Ctrl. All X South		7.381+		-25.088
		(4.178)		(25.929)
Gov.+ UH X South		-1.239		-19.244**
		(2.071)		(3.032)
Gov.+ LH X South		-6.801*		43.992+
		(2.755)		(22.645)
Governor X Phys. Dom.			-.096	-.107
			(.108)	(.116)
Upper Hse X Phys. Dom.			-.266	.093
			(.392)	(.188)
Lower Hse X Phys. Dom.			-1.489+	1.654*
			(.873)	(.705)
State. Leg. X Phys. Dom.			1.706	-1.704*
			(1.089)	(.760)
Ctrl. All X Phys. Dom.			-.649	2.910**
			(1.654)	(1.110)
Gov.+ UH X Phys. Dom.			.465	.100
			(.424)	(.285)
Gov.+ LH X Phys. Dom.			.584	-2.876**
			(1.499)	(1.058)
Governor X Phys. Dom. X South				.781**
				(.192)
Upper Hse X Phys. Dom. X South				-16.252**
				(2.828)
Lower Hse X Phys. Dom. X South				60.523*
				(25.610)
State. Leg. X Phys. Dom. X South				-45.920
				(27.957)
Ctrl. All X Phys. Dom. X South				47.355
				(28.893)
Gov.+ UH X Phys. Dom. X South				14.861**
				(2.612)
Gov.+ LH X Phys. Dom. X South				-60.689*
				(26.488)
Number of Obs.	550	550	550	550
Number of Clusters	50	50	50	50
R^2	.87	.89	.88	.91
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

(1.19)

Table 1.20

Determinants of Regulations: Full Specifications *Dep. Var. is # Backlash Regulations*				
	(1)	(2)	(3)	(4)
Dem. Ctrls.	Yes	Yes	Yes	Yes
Dem. Ctrls. X South	No	Yes	No	Yes
Dem. Ctrls. X Phys. Dom.	No	No	Yes	Yes
Dem. Ctrls. X Phys. Dom. X South	No	No	No	Yes
Number of Obs.	550	550	550	550
Number of Clusters	50	50	50	50
R^2	.87	.89	.88	.91
P-value All Regressors are Zero	.41	.00	.15	0
P-value Phys. Dom. Intracts. are Zero			.07	0
StateFE	Yes	Yes	Yes	Yes
YearFE	Yes	Yes	Yes	Yes

(1.20)

Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on health expenditures and GSP from CMS. Data on Democratic control obtained from the Statistical Abstract of the United States. Data on physician dominance of health interest groups (fraction of health lobby organizations by primary care clinic organizations obtained as personal communication from Virginia Gray.

Table 1.21

Determinants of Regulations: Demonstration *Dep. Var. is # Backlash Regulations*							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dem. Ctrl. Type	Dem. Gov.	Dem. U. Hse	Dem. L. Hse	Dem. Gov. + UH	Dem. Gov. + LH	Dem. State. Leg.	Dem. Cntrl All
Dem. Cntrl.	.263	.064	.202	.027	.325	.063	.092
	(.236)	(.219)	(.246)	(.233)	(.397)	(.239)	(.373)
Phys Dom. X Dem. Cntrl.	-.026	.024	.005	.068	.042	.050	.082
	(.055)	(.071)	(.101)	(.076)	(.183)	(.092)	(.154)
Dem. Cntrl. X South	-.226	-.221	.027	-.798	-1.074	-.252	-1.053
	(.680)	(.416)	(.493)	(.858)	(.918)	(.491)	(.995)
Phys Dom. X Dem. Cntrl. X South	.177	.075	-.070	.266	.423	-.040	.428
	(.181)	(.188)	(.479)	(.238)	(.400)	(.469)	(.396)
Number of Obs.	550	550	550	550	550	550	550
Number of Clusters	50	50	50	50	50	50	50
R^2	.86	.86	.86	.86	.86	.86	.86
StateFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YearFE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

(1.21)

Standard errors clustered by state in parentheses. Data on state regulation of managed care obtained from the NCSL. The regulations variable is the sum of all regulations in force in the given state and year. Data on the percentage of state population enrolled in HMOs in 1995 is obtained from the Statistical Abstract. Data on health expenditures and GSP from CMS. Data on Democratic control obtained from the Statistical Abstract of the United States. Data on physician dominance of health interest groups (fraction of health lobby organizations by primary care clinic organizations obtained as personal communication from Virginia Gray.

Chapter 2

Economic Discontinuities at Borders: Evidence from Satellite Data on Lights at Night

2.1 Introduction

Does political economy matter for economic growth? If yes, what are the channels through which it matters? A critical difficulty in answering this question is the endogeneity of political economy and politically determined variables such as institutions, public goods provision, macroeconomic policies, education and others: they may be correlated with unobserved variables that also affect growth. In particular, they may be correlated with geographic variation: countries that are more favorably endowed by geography may have better institutions and a better-functioning government. A large and fruitful literature has endeavored to resolve the endogeneity problem by using instrumental variables (La Porta et al. (1998), Acemoglu et al. (2001)) and found large effects of institutions.

This paper presents an alternative approach to measuring the impact of political economy, as opposed to geography, on growth: exploiting spatial discontinuities created by national borders. While borders are obviously determined endogenously (e.g. through war or national reunification), their precise location is often arbitrary, following a river or a line of latitude or longitude, and without regard to the characteristics of localities within 30 or 50 kilometers of the proposed line. Therefore, it may be expected that nearby locations separated by national borders should be similar in terms of geography and other local variables, but different in terms of national-level variables including political economy. Moreover, if these localities are small enough, it is overwhelmingly likely that while they are affected by national-level variables of the country that they are part of, they do not affect these variables themselves. Hence, we can view locations near the border as subjected to a natural experiment, in which they are randomly assigned to different national-level institutions, and in particular, to political economy.

While many potential determinants of growth change discontinuously at national borders, I argue that border discontinuities can be used to assess the impact of political economy on economic activity because these determinants are produced by government activity. Institutions, such as existence of the rule of law, protection of property rights and political freedom, are perhaps a classic example of such a "spillover" determinant of growth.[1] Some public goods, such as a nationwide infrastructure grid, or a program of universal education provision, may also exhibit such spillovers through decreasing transaction costs and creating human capital externalities. Finally, culture and the level of trust may be affected by national-level shocks through centralized television programming and the presence of a common language. However, all of the spillover effects described above are mediated by the one economically relevant variable that necessarily changes at a national border: the identity of the national government. Governments choose (or perpetuate) institutions of private property and manage the national stocks of public goods. To the extent that culture is affected by national-level shocks, it tends to be shaped by the actions of the government, such as the creation of a government television or radio channel that is accessible in all parts of the country, or a policy of cultural and linguistic homogenization. Therefore, discontinuities in economic activity at borders should be interpreted as estimates of the importance of government activity (short-run policies or long-run institutions and culture) for the level and growth of the economy, for good or for ill. The presence of large border discontinuities in economic activity suggests that actions taken by governments (again, over the short or the long run) have powerful effects on income per capita, whereas their absence indicates that economic shocks, rather than political actions, account for differences in the wealth of nations.

An alternative view of borders could be that they are discontinuities in the level of transaction costs in the purely private economy. The fact that language and culture are different on different sides of borders, as well as the existence of explicit tariffs and subsidies that discourage trade, imply that the economies on the two sides of the border are less connected by trade than two economically comparable regions within the same country. Therefore, a shock to the private economy in a region of one country (such as a poor harvest, an influx of immigrants, a commodity bubble, a discovery of natural resources or a labor dispute) that spills over into neighboring regions in terms of changes in wages, prices and demand, may fail to spill across a national border because of the discontinuous increase in transaction costs along that border. Therefore, border discontinuities in economic activity may exist without being mediated by government activity. However, this explanation relies entirely on trade as a transmission channel of economic disturbances, and therefore, is easy to confirm or rule out. In Section 2.5, I present evidence that the (normalized) amount of trade between two countries does not explain the size of the economic discontinuity at their border or its relationship to key variables that are determined by governments.

To compute GDP and growth in narrow bands around national borders, I use satellite data on lights at night collected by the Earth Observation Group (NOAA) in the DMSP-OLS satellite program. The night lights dataset has been first described by Henderson et al. (2011), who have shown a strong correlation between the amount of light emitted from a country and its GDP, both for levels as well as for growth rates. Night lights are an ideal and indispensable data source for this project because they are one of the few indicators of economic activity that exist at a sufficiently fine resolution to allow the analysis of narrow neighborhoods around national borders, as well as because the method of their collection is continuous across national borders. National accounts data is fundamentally unsuitable for such a project because it is available, at best, only at a regional level, and does not permit considering regions other than large political subdivisions, thus making it inappropriate for a regression discontinuity analysis. Survey data may overcome this problem as it samples individuals or villages rather than geographical units, but it would introduce an artificial discontinuity at borders because each survey is conducted within a

[1] In developing countries where the central government is weak, institutions may be nonuniform across a country, as hypothesized by Acemoglu and Dell [2010]

single country. Hence, errors attributable to the questionnaire or to the performance of different survey teams would be different on different sides of national borders. On the other hand, even if there is important spatial heterogeneity in the way that satellites record lights data, the remote sensing process should be continuous, and therefore, roughly stable within a neighborhood around a given border.[2]

This paper is part of a large literature on the effects of political economy on growth that includes Dell (2010) on the impact of colonial forced labor in Peru, Banerjee and Iyer (2005) on the present-day influence of taxation systems in British India, Nunn (2008) on the persistent effects of the African slave trade, and Larreguy (2011) on the persistence of colonial institutions in Nigeria. This paper is closest to Michalopoulos and Papaioannou (2011a), who use night lights data to look at the (non)importance of institutions for economic activity within African ethnicities split by national borders. The innovation of this paper is 1) its much wider scope in considering the universe of borders around the globe rather than an institution in a particular country, 2) its use of regression discontinuity neighborhoods rather than existing regions, and 3) its development of an econometric theory to accommodate the special nature of the data used.

Using the amount of lights per capita as a proxy for economic activity around borders, I document a strong and highly significant relationship between national GDP and GDP at the border. As one moves from a poorer to a richer country sharing a border, the amount of light per capita (calibrated to be comparable to GDP per capita) rises on average by 40 log points (50%). Moreover, for every 1% difference in GDP per capita between the two bordering countries, there is a 0.63% difference between the amount of light per capita at their borders. More surprisingly, there also exists a relationship between differentials in growth of lights per capita across a border and differences in growth in the bordering countries over a 20-year period from 1990 to 2010.[3] As one moves from a slower-growing to a faster-growing country, the 20-year growth rate of light per capita rises on average by 2.6 percentage points, and for every 1 percentage point difference in the growth rates of GDP per capita of two bordering countries, there is a 0.88 percentage point difference in the growth rate of lights per capita of these countries at their mutual border.This finding is unexpected because while differentials in levels of income (the world distribution of income) tends to be persistent, growth rates are much more volatile, both across time and within a single country. Therefore, an association between differences in national growth rates and differences in border growth rates suggests that border discontinuities represent not only accumulated effects of large historical events in the past, but that they represent factors that promote or stymie current economic activity and the ability of people to take advantage of or overcome their past. If discontinuities at the border can be attributed to political economy, this finding shows a substantial effect of suitable government activity for growth in a country over a short period of time.

The finding can be highlighted in two pictures, both based on Elvidge (2003). The first, Panel 1 in Figure 2.1 is a satellite photo of North Korea and South Korea, the former covered in darkness, the latter lit up, with the light beginning right at their common boundary. The second, Panel 2, is a comparison of two satellite photos of Ukraine and its neighbors: the first taken in 1992 and the second taken in 2000. Areas that gained light are represented in white, whereas areas that lost light are represented in black. The comparison reflects the obvious fact that during the

[2]The recorded brightness of lights may depend on cloud cover, humidity and other atmospheric conditions in a region, but it is implausible that national borders consistently conform to atmospheric fronts. Robustness checks with controls for temprerature, precipitation, altitude and slope on both sides of borders do not alter the results.

[3]Strictly speaking, this is an 18-year growth rate because the lights data does not start until 1992. However, high-resolution population data is not available for 1992 but is available for 1990, so that is the data I use to compute per capita growth rates. I will refer to this measure as a 20-year growth rate throughout.

transition from communism in the 1990s, Ukraine (and its ex-Soviet southern neighbor Moldova) contracted much more severely than did Poland, Romania and Hungary, which had exceeded their 1992 GDP by 2000. What is striking about the picture is that there is a border discontinuity in growth in nighttime lights – there are very few white dots in Ukraine and Moldova and very few black dots in Poland, Hungary and Romania (except for some in the Carpathian mountains, which are not on a border). Almost all places in the former Soviet republics had contracted, and almost all places in Eastern European nations had expanded between 1992 and 2000, even those very close to the borders between these two sets of countries. Particularly striking is that the westernmost tip of Ukraine (Ruthenia) had been part of Ukraine for only 50 years before the time period in question and had been in a political union with Hungary, Slovakia and Romania for most of its prior history[4], and yet, it experienced a decline in lights as did the rest of Ukraine, while the neighboring parts of these countries experienced growth in lights. Moreover, Moldova and Romania (southwest corner of picture) share the same language, religion and culture (although they have been politically separate for most of their modern history), but have had radically different growth experiences in the 1990s with Romania growing, Moldova shrinking, and the growth experience changing discontinuously at the border. Figure 2.2 formalizes Figure 2.1 by presenting local average lights per capita (and growth in local average lights per capita) for the places described. We can see very strong and very clean discontinuities at borders in all three graphs. My finding in this paper is that these pictures are not anomalies, but rather very stark depictions of a general pattern.

It is also intuitive that border discontinuities present lower bounds for the importance of government activity for the economy. First, borders are porous, which means that trade and migration may mitigate differences created by government activity on the different sides of the border. Second, and more fundamentally, border discontinuities are biased downward in the night lights dataset because of blooming: satellite-recorded light tends to spread away from its source, thus leading light generated on one side of the border to be seen on the other side of the border. In Section 2.5, I document that poorer countries tend to experience a rise in lights per capita relative to richer countries as one approaches their mutual border.

The above results are most straightforwardly obtained by considering lights in narrow neighborhoods around borders, which is a version of local constant regression discontinuity estimation and is known to be biased (with the bias going to zero asymptotically) if the derivative of the outcome variable with respect to the running variable at the border is large (if lights per capita converge rapidly very close to the border). Regression discontinuity estimates with better bias behavior can be obtained by using local polynomial estimation. A complication in using local polynomial estimation to calculate border discontinuities with nighttime lights data is that the data generating process does not obey the standard assumption of independently generated data with the number of observations at each site going to infinity. Instead, the nighttime lights data constitute a global census of visible nighttime lights, taken at a fixed resolution. Therefore, the asymptotics for the regression discontinuity estimator must be calculated as the resolution of the data goes to infinity (the pixel size goes to zero) rather than as the domain of the pixels expands to infinity. Such an asymptotic scheme is referred to as *infill asymptotics* in the spatial econometrics literature. A natural assumption for such data is that the errors from trend of neighboring data points are correlated. A contribution of this paper is to derive the properties of the local poly-

[4]Ruthenia had been a part of the Kingdom of Hungary since 1526, and of the Habsburg empire (which included Hungary, Slovakia, and the parts of Poland and Romania that are visible in this picture) since 1699. After the collapse of the Habsburg empire as a result of World War I, Ruthenia became a part of Czechoslovakia in 1918. Ruthenia was annexed by the Soviet Union in 1945 as a result of World War II, and attached to the Ukrainian Socialist Soviet Republic, which became the independent country of Ukraine in 1991.

nomial estimator under infill asymptotics with correlated errors. I prove that with very general assumptions on the covariance structure of the outcome variable the local polynomial estimator is consistent, and has a smaller asymptotic variance than it would if the errors from trend were independent.[5] Intuitively, the local polynomial estimator exploits the correlation in the errors, so that only their unpredictable component contributes to the asymptotic variance. In the special case that the error from trend is mean-square continuous (has no unpredictable component), the local polynomial estimator converges at a nonstandard rate of $1/\sqrt{h}$, where h is its bandwidth.

A further complication of using local linear regression is contamination of the night lights dataset. It is well known (Doll 2008) that the satellites recording nighttime light density tend to attribute light generated at a particular site to nearby sites as well. For example, the Pacific Ocean is lit up as far as 50 kilometers away from the California shore near Los Angeles. This phenomenon is known as overglow. While local linear regression is very important for recording the potential narrowing of differences in economic activity at borders because of cross-border trade, it also will pick up the convergence of nighttime lights density at borders because of overglow, which will complicate finding any discontinuities in economic activity that may exist. In this paper, I propose a novel correction for overglow by calibrating an overglow function over territories on the borders of wastelands and using this function to correct nighttime lights values at borders. I implement this correction to improve my local linear estimates and demonstrate that while overglow can be a substantial problem for local linear analysis, it ceases to be a problem once the correction is made.

I then go beyond providing evidence of discontinuities at national borders, and hence of the importance of political economy to economic activity, and attempt to uncover which politically determined variables are useful in understanding and explaining border discontinuities. First, I show that richer sides of borders do not tend to have more public goods – specifically, roads, railroads and utilities – in narrow neighborhoods of the border than poorer sides of borders do. Therefore, one cannot explain border discontinuities through differences in local public good provision. However, public goods provision could still explain border discontinuities if it has large spillovers – for instance, good infrastructure in the country as a whole may benefit a region with worse infrastructure through endowing the region with richer trading partners from other regions.

Restricting myself to analyzing discontinuities in growth rates, I perform a correlational analysis to see whether they can be explained by several national-level variables frequently discussed in the cross-country growth literature. I show that when the extent of the World Bank measure of the rule of law (which considers the impartiality of the judicial system, the quality of contract enforcement, and the protection of private property against confiscation) is accounted for, the correlation between differences in national growth and differences in growth at the border falls substantially and becomes insignificant, while differences in the rule of law between two countries are associated with higher differences in subsequent growth at their borders. While there is similarly an association between differences in countries' initial levels of public goods provision (proxied by the fraction of roads paved) and differences in their subsequent growth rates at the border, controlling for public goods provision does not eliminate the association between differences in national growth and differences in growth at the border. Accounting both for the rule of law and for public goods provision, the association between differences in the rule of law and differences in border growth remains intact, while the association between differences in public goods and differences in border growth shrinks substantially. I further show that the rule of law retains its explanatory power when I control for contracting institutions (Acemoglu and Johnson [2005]), political freedom, the average amount of education, and a measure of interpersonal trust from the World Values Survey.

[5]This result is most closely related to Card and Lee (2008)

The paper is organized as follows: Section 2.2 describes the data. Section 2.3 discusses the efficiency improvement of the local polynomial estimator in an infill asymptotics setting with correlated errors. Section 2.4 provides baseline results on border discontinuities in GDP and growth rates as robustness checks to accounting for local climate and public goods variation. Section 2.5 explores the role of property rights protection in generating border discontinuities and provides evidence that border discontinuities do not arise because of the discontinuous barriers to trade that borders pose. Section 2.7 concludes.

2.2 Description of the Data

2.2.1 The Night Lights Dataset

Data on light radiance at night is collected by the DMSP-OLS satellite program and is maintained and processed by the Earth Observation Group and the NOAA National Geophysical Data Center. Satellites orbit the Earth every day between 20:30 and 22:00, sending images of every location between 65 degrees south latitude and 65 degrees north latitude at a resolution of 30 arc-seconds (approximately 1 square km at the equator). The images are processed to remove cloud cover, snow and ephemeral lights (such as forest fires and gas flaring) to produce the final product available for download at

http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html

Each pixel (1 square kilometer) in the radiance data is assigned a digital number (DN) representing its radiance. The DNs are integers ranging from 0 to 63, with the relationship between DN and radiance being

$$\text{Radiance} \propto \text{DN}^{3/2}$$

(Elvidge et al. 1999). However, pixels with DN equal to 0 or 63 may be top- or bottom-censored. Another known problem with the lights data is the presence of overglow and blooming: light tends to travel to pixels outside of those in which it originates, and light tends to be magnified over certain terrain types such as water and snow cover. All of these problems tend to make nearby pixels more similarly lit than they should be, thus working against the hypothesis of this paper.

The night lights dataset has been extensively analyzed in the remote sensing literature for its utility in predicting economic activity; see Elvidge et al. (1997), Sutton et al. (2007), Doll (2006), Ghosh et al. (2010) and Elvidge et al. (2012). Baugh et al. (2009) thoroughly describes the construction of the night lights dataset, and Doll (2008) comprehensively discusses its uses and pitfalls. Its pioneering use in the economics literature has been Henderson et al. (2011). Chen and Nordhaus (2010) discuss the limitations of the lights dataset; in particular, they argue that the relationship between light density and output density becomes uninformative because of top-censoring and bottom-censoring at $DN = 63$ and $DN = 0$. Michalopoulos and Papaioannou (2011a and b) use the night lights dataset to construct a proxy for output per capita in African ethnic territories to assess the consequences of partitioning ethnicities during the Scramble for Africa.

2.2.2 Gridded Population of the World Data

The Gridded Population of the World (GPW) dataset is constructed and maintained by the Socioeconomic Data and Applications Center (SEDAC) at the Center for International Earth Science Information Network at the Earth Institute at Columbia University. The dataset compiles population information from national censuses for very small political units (municipalities, census tracts) in order to achieve its resolution. Within a political unit, population is distributed uniformly.

2.2.3 Other Data

In Sections 2.4 and 2.6 I use a number of geographic, political and social variables to validate the regression discontinuity and explore correlations between discontinuities in growth rates at national borders and politically affected determinants of economic growth for the bordering countries. I obtain data on the distributions of temperature and precipitation levels (means, medians, annual and mean daily ranges, coefficients of variation) at 30 arcsecond resolution from the dataset constructed by Hijmans et al. (2005) and distributed on the WorldClimate website: http://www.worldclim.org/. I obtain elevation at 30 arcsecond resolution from the USGS Shuttle Radar Topography Mission (SRTM), and use the data to compute slope in ArcGIS. Data on roads, railroads and utilities is from the US National Imagery and Mapping Agency, and originally from the Digital Chart of the World. I also use several country-wide covariates from standard sources in the literature. I obtain data on the presence of the rule of law and other governance indicators from the World Governance Database (WGI) sponsored by the World Bank.[6] Data on the fraction of roads paved and on the amount of time required to enforce a contract is obtained from the World Bank's World Development Indicators. Religious composition and legal origin of countries is obtained from La Porta et al. (1998). Political freedom is measured using the Freedom House Political Rights Index, from Acemoglu, Johnson, Robinson and Yared (2008). Average years of education are obtained from Barro and Lee (2010). A measure of trust is obtained from the World Values Survey via La Porta (2011). Bilateral national-level trade data is obtained from the IMF, Direction of Trade Statistics. Finally, I use national GDP data from the Penn World Tables, Mark 7.1.

2.3 Regression Discontinuity under Infill Asymptotics

2.3.1 Discussion of Literature

The methodology for regression discontinuity has been extensively developed by Hahn, Todd and van der Klaauw (1999), Porter (2003) and Card and Lee (2008) and is reviewed in Imbens and Lemieux (2008) and Lee and Lemieux (2010). However, the asymptotics in these papers assume either that observations are independent or that the diameter of the domain from which observations are drawn expands to infinity as the number of observations tends to infinity. In the context of estimating border discontinuities from nighttime lights data, these assumptions are unsuitable because the nighttime lights represent a regular grid of observations in a fixed domain (a neighborhood of the border in question). Moreover, nearby values of lights are very likely to be correlated. In particular, it makes sense to think of asymptotics in the nighttime lights dataset as an improvement in the resolution of the regular grid rather than as an increase in the number of observations. Such an asymptotic analysis, while very uncommon in econometrics, is frequently performed in spatial statistics and geology, and is referred to as *infill asymptotics.* One contribution

[6] In results not reported, I also use average protection from expropriation risk between 1985 and 1995 from Political Risk Services via Acemoglu, Johnson and Robinson (2001), and an index of property rights protection from the Economic Freedom of the World database (variable 2C).

of this paper is to develop the asymptotic properties of the local polynomial estimator in an infill asymptotics setting.

There is an emerging literature on the econometrics of processes defined on two-dimensional surfaces rather than on a time axis (Conley 1999, Hansen et al. 2008, Robinson 2011), which extends insights from the time series literature in multiple dimensions. However, this literature concerns itself entirely with increasing-domain asymptotics: it is assumed that as the number of observations increases, the domain of the grid tends to infinity. There is substantial discussion of processes under infill asymptotics in the spatial statistics literature, most notably by Stein (1987, 1999), particularly in the context of kriging, or spatial interpolation and extrapolation. Stein's results are derived for covariance stationary Gaussian processes when the statistician has substantial prior information about the shape of the covariance function, such as the functional form. The results presented here will be valid for substantially more general processes without any functional form assumptions, and without the assumption of covariance stationarity. General references on spatial statistics are Cressie (1993) and Schabenberger and Gotway (2005).

This section is closest to Card and Lee (2008), who consider regression discontinuity estimation when the running variable is observed at a number of discrete sites, and perform asymptotic analysis as the number of these sites goes to infinity. However, Card and Lee (2008) assume that the errors between the assumed and the true functional forms of the relationship between the outcome variable and the running variable are independent. The analysis in this paper will relax this assumption, which will entail a substantially different analysis from that of Card and Lee.

2.3.2 Properties of the Local Polynomial Estimator under Infill Asymptotics

I consider the properties of the standard local polynomial estimator computed for an outcome variable y that is observed on a regular one-dimensional grid;[7] hence, for the sequence $\left\{y\left(\frac{u}{N}\right)\right\}_{u=1}^{N}$. As N goes to infinity, it is clear that y is never observed outside of $[0,1]$, but it is observed at an increasing frequency. The running variable x is distance: $x = \left\{\left(\frac{u}{N}\right)\right\}_{u=1}^{N}$The core result is that for estimation of a regression discontinuity at a single point, the local polynomial estimator is consistent, and its asymptotic variance is smaller than the probability limit of the traditional White estimator for heteroskedasticity based on the residuals of the local polynomial estimator. The intuition for this fact is that when the errors from the deterministic relationship between the outcome and the running variable are correlated, the weighting scheme of the local polynomial estimator exploits this correlation to predict the outcome at the discontinuity, which makes the effective magnitude of the error equal to that component of it that is unpredictable. The White estimator, however, assumes that all errors are independent and computes the variance accordingly. Therefore, when facing infill data with correlated errors, the typical variance estimator is overly conservative.[8]

I further present an estimator that is consistent for the true asymptotic variance of the local polynomial estimator. Instead of being based on the squared residuals, it is based on squared differences of residuals from adjacent observations. This estimator filters out both the deterministic

[7]The mathematical results for a two-dimensional grid are straightforward extensions of the results presented in the Appendix. The empirical results are also very similar, but computationally more difficult to obtain.

[8]The only substantive assumption necessary to prove this is that the process of error terms can be decomposed into the sum of a process of independent random variables and a process of correlated random variables whose covariance function is sufficiently smooth. If there is no unpredictable component to the error term (the error process is mean-square continuous), the local polynomial estimator converges to its probability limit at a nonstandard rate of $1/\sqrt{h}$.

trend and the correlated component of the residual, regardless of their functional form and correlation structure, leaving only the idiosyncratic variability at each site to contribute to the estimated variance. Hence, the variability in the correlated component of the error term is not (mistakenly) attributed to the estimator.

The variance estimator I propose is given by

$$\hat{V}_{1,N} = \frac{1}{2} e_1' D_N^{-1} \left[\sum_{u=1}^{N} \hat{e}_{N,u}^2 \left(\frac{u}{N} \right) \frac{1}{Nh} k^2 \left(\frac{u}{Nh} \right) X \left(\frac{u}{Nh} \right) X \left(\frac{u}{Nh} \right)' \right] D_N^{-1} e_1 \tag{2.1}$$

where $k()$ is a kernel, h is the bandwidth, N is the number of observations in the fixed interval under consideration, $X()$ is a vector of polynomials in distance to the border, D_N is the denominator of the local polynomial estimator, e_1 is a vector with first component equal to 1 and all the others equal to zero, and $\hat{e}_{N,u}\left(\frac{u}{N}\right) = \hat{e}\left(\frac{u}{N}\right) - \hat{e}\left(\frac{u-1}{N}\right)$ is the difference between adjacent residuals obtained from local polynomial estimation. One should contrast this variance estimator with the traditional White estimator, which is given by

$$\hat{V}_N^{OLS} = e_1 D_N^{-1} \left(\sum_{u=1}^{N} \hat{e}^2 \left(\frac{u}{N} \right) \frac{1}{Nh} k^2 \left(\frac{u}{Nh} \right) X \left(\frac{u}{Nh} \right) X \left(\frac{u}{Nh} \right)' \right) D_N^{-1} e_1$$

It is also useful to note the relationship between the proposed variance estimator $\hat{V}_{1,N}$ to the classical estimator of the variogram proposed by Matheron (1962) (see also Schabenberger and Gotway 2005). Matheron's estimator of the variogram $\gamma(\tau) = E\left((e(u) - e(u-\tau))^2\right)$ is given by

$$S(\tau) = \frac{1}{|A(\tau)|} \sum_{A(\tau)} (e(s) - e(s+\tau))^2$$

where $A(\tau)$ is the set of pairs of points in the space that are τ apart.

The proposed estimator $\hat{V}_{1,N}$ can be thought of as estimating the limit of the variogram as τ goes to zero. The kernel density estimation ensures that the limit is computed for the variogram that holds at $x = 0$, and therefore allows for the variogram to be nonstationary and change over regions of space.

The estimator $\hat{V}_{1,N}$ converges in probability to a smaller value than the estimator $\hat{V}_N^{OLS}$, but in finite samples, $\hat{V}_N^{OLS}$ may be numerically smaller. Therefore, I use the minimum of the two estimators when computing the variance. I also consider data at a resolution of 1 km (the resolution at which the night lights data is available), which brings $\hat{V}_{1,N}$ closer to its probability limit. All propositions and proofs are relegated to the Appendix.

2.4 Baseline Results

2.4.1 Calibration

To convert lights data into a single quantity comparable to GDP, I assume a low-parameter function approximating the relationship between DN and output density, and calibrate its parameters using aggregate light density for countries and national accounts data on GDP per capita. Specifically, I estimate the parameters of the function using nonlinear least squares, in which I try to explain GDP density per unit area in a country with measures of light density for the country constructed using pixel digital numbers. The assumed relationship is

$$\ln(1+y_i) = c + \ln\left(1 + c_0 * v_{0,i} + c_b * \sum_j \left(j^d\right) * v_{j,i} + c_1 * v_{63,i}\right) + \varepsilon_i \tag{2.2}$$

where i indexes countries, y_i is GDP density of country i (obtained from the World Bank), $v_{j,i}$ is the fraction of pixels with digital number equal to j in country i, and ε_i is the error term. I use the transfer function $\ln(1+x)$ rather than $\ln(x)$ because the latter is not defined for $x = 0$ while the former is defined for all nonnegative x (and some negative values of x as well). This is not a problem in the estimation of the calibration equation, as the light density of no country is equal to zero; however, the light density at some borders does attain the value zero, which explains the need to use such a specification. For reasonable values of the output density y_i, $\ln(1+y_i)$ is indistinguishable from $\ln(y_i)$, while the parameters on the right hand-side allow full parametrization of the scale of the index of fractions of pixels relative to 1. Note that the measure of light density used by Henderson et al. (2011) would be equivalent to setting

$$c_0 = c_1 = 0, d = 1$$

and the measure used by Chen and Nordhaus (2010) would amount to

$$c_0 = c_1 = 0, d = 3/2$$

(The values c and c_b would be set sufficiently large in magnitude to make the 1 in the parentheses inconsequential). I estimate equation (2.2) for every satellite-year in the DMSP-OLS dataset, using the Chen and Nordhaus (2010) specification as my initial values. For multiple years (in particular 2000 and 2005), the estimates of the top-censoring and bottom-censoring coefficients c_0 and c_1 are equal to zero, suggesting that top-censoring and bottom-censoring is not a particularly important limitation of the data.

2.4.2 Descriptive Analysis and Graphs

Computation of the Dependent Variable

Since national borders tend to differ substantially in length, I standardize them by dividing each border into pieces corresponding to its intersection with a 1-degree by 1-degree grid superimposed on the world map. I obtain 1352 border pieces, having started with 270 borders. In all my computations of standard errors, I cluster the standard errors by border (rather than border piece) so I have 270 clusters. Throughout the rest of the paper, I will refer to border pieces as borders to minimize terminology unless I need to make the distinction explicit.

To obtain a lights-based proxy for economic activity around a given border for a given year, I construct neighborhoods containing all points whose shortest distance to the border is less than X kilometers, and use an ArcGIS Python program to compute the fraction of pixels with each digital number for each side of the border within the neighborhood. I then use the calibrated values of c, c_0, c_1, c_b and d to compute the right-hand side of equation (2.2) for each side of the border, thus obtaining a proxy for the output density of the given country within X kilometers of the given border. Finally, I multiply by the area and divide by the population of this region (as discussed in Section 2.2, I obtain population at 2.5 arcminute resolution from the Gridded Population of the World dataset) to obtain a lights-based estimate of GDP per capita for the given country within X kilometers of the given border.

There are good reasons to expect that this calibration procedure creates a variable that is, on average, close to the true value of GDP per capita in the region of interest. Henderson et al. (2011) and Chen and Nordhaus (2010) document the tight association between GDP density and measures of light density that are similar to the one used in this paper (in fact, they are special cases of my measure). The fit of the selected specifications to the GDP data for countries is very good as the first panel of Figure 2.3 attests: the lights explain approximately 73% of the variation in GDP per capita, and the plot looks approximately linear. I also present the relationship between the growth of the calibrated lights series and the growth of GDP per capita in the second panel of Figure 2.3. The fit is not as good (the lights explain only 23% of the variation in growth), but still quite strong, and the positive correlation is unmistakeable.

Descriptive Statistics

Table 2.1 provides descriptive statistics for the main variables of interest: log lights per capita at borders and their growth rate, log lights per capita and log GDP per capita nationwide, with their growth rates, and some covariates related to institutions and public goods. I present the mean and the standard deviation of each variable, as well as the mean and the standard deviation of each variable computed over the richer (or higher-growing) and poorer (or lower-growing) sides of borders exclusively. The descriptive statistics foreshadow more formal results. We immediately see that log lights per capita are higher on richer sides of borders than on poorer ones, and that the mean difference between the two is about half of the mean difference between nationwide log lights per capita (or nationwide log GDP per capita) of the bordering countries. We see even starker results for differences in growth rates of light per capita between higher-growing and lower-growing countries at their mutual border. In panel 2 of the table, where we look at institutions and public goods, we see that the rule of law of the higher-growing country at a border are much better on average than the rule of law of the lower-growing country at that border. Public goods (proxied by the fraction of roads paved) are also better in the higher-growing country at a border, but not by much, as is trust and the number of years of schooling. Interestingly, local public goods (log roads near the border) are very close to each other on the higher-growing and lower-growing sides of borders on average.

Elementary Discontinuity Plots and Correlations

I now present several elementary graphs that suggest large discontinuities in GDP per capita and its growth rate at national borders. Panel 1 of Figure 2.4 shows a discontinuity plot of lights per capita (predicted GDP per capita using lights) against distance from border in the direction of the richer country at the border as predicted by the lights calibration. To construct this plot, I identify the richer country and the poorer country at each border according to which one has the higher lights per capita. I then pool all points over the portions of the discontinuity neighborhoods that belong to the poorer countries, and compute the average lights per capita for 5-km intervals of distance to the border. I repeat the same procedure for the richer countries and plot the averages as a function of distance to the border either for the richer (on the right) or for the poorer (on the left) side.

It is apparent that there is a discontinuity at the border crossing point, with the richer (according to lights) side of the border having a GDP per capita at least 0.2 log points (about 22%) higher than the poorer side. The last point on the poorer side (at -5 km) is approaching the points on the richer side, but the other points on the poorer side are far removed from those on the richer side (by at least the 0.2 of the discontinuity). This can be explained by overglow in the data: light from the richer side of the border illuminates the poorer side, making it appear to be richer.

Panel 2 of Figure 2.4 presents the same discontinuity plot, but with countries categorized as rich or poor on the basis of GDP per capita from the Penn World Tables rather than on the basis of national lights per capita. Constructing this plot has both advantages and disadvantages compared with Panel 1: I use a real rather than calibrated output measure in this plot, but I also have to compare lights to GDP in this plot, which introduces measurement error. We again see a clear discontinuity at zero that is substantially larger than any other difference between consecutive points, with a somewhat smaller magnitude than before.

The discontinuity plots in Figure 2.4 elide the fact that countries are extremely heterogeneous and the difference in economic activity at the border between a poorer and a richer country may vary widely, even if it is large and positive on average. One would instead expect the difference at the border to be somehow related to the difference in economic activity between the two countries overall: countries with wide disparities in income per capita (like North Korea and South Korea) should have larger discontinuities at their common border than countries with similar incomes (like France and Germany). Panel 3 of Figure 2.4 presents a plot of differences in log lights per capita at a border against differences in log lights per capita in the bordering countries (each border difference being weighted by population). The positive correlation and its strength are manifest. Panel 4 shows that if differences in log lights per capita at the border are plotted against national differences in log GDP per capita (from the World Bank), the correlation is similar.

It is perhaps not very surprising that GDP per capita may be discontinuous at national borders since borders change infrequently, and in many cases remain stable for decades and even centuries, allowing a discontinuity to accumulate. However, just as there is a discontinuity in GDP per capita across borders, so there is one in GDP growth over relatively short periods of time. Panel 1 in Figure 2.5 presents a discontinuity plot similar to Panel 1 in Figure 2.4, but computing the average annualized 20-year growth rate in lights rather than the average amount of lights at each location, and comparing countries with higher growth in national lights per capita with countries with lower growth in national lights per capita rather than richer countries with poorer countries.

There is an extremely prominent discontinuity in growth of lights per capita at the border, with the higher-growing country (according to lights) growing by over 1 percentage point more each year for 20 years, on average, than the lower-growing country. Thus, not only have borders contributed to (static) discontinuities in levels over the decades that they remain unchanged, but they also create (dynamic) discontinuities in growth rates over time periods as short as 13 years. Panel 2 of Figure 2.5 shows the same discontinuity plot with countries categorized as higher-growing or lower-growing based on national GDP per capita growth. The discontinuity is again prominent, but more modest in magnitude. Panels 3 and 4 show correlation plots between differences in growth at the border and differences in national growth in output per capita (measured with lights or GDP) similar to the same panels of Figure 2.4: the conclusions are the same.

2.4.3 Overglow Correction

A problem in the analysis of nighttime lights data, which is particularly severe for its use in a regression discontinuity design, is the presence of oveglow and blooming. The satellite sensor tends to observe light in territories that are devoid of human activity, but are close to densely settled areas. For example, Doll (2006) presents an illustration of light from Los Angeles being observed over the Pacific Ocean as far as 50 kilometers away the California shore. In a regression discontinuity analysis, overglow operates continuously across borders and transforms any discontinuities into nonlinearities. In Figure 2.2, we saw the profound effect of overglow at the border between the two Koreas, and in Figures 2.4 and 2.5 we see the discontinuity gaps shrinking as one approaches the border.

To alleviate the impact of overglow, I construct a parsimonious model for the overglow process, estimate the parameters of the model, and use it to correct my measures of nighttime light density. I assume a one-dimensional autoregressive model (to parallel my approach to estimating border discontinuities) in which light observed over one 10-km strip of land increases the amount of light observed in neighboring 10-km strips (the width again chosen to parallel the border discontinuity setting). Specifically, if I divide a tract of territory into square sites, subdivide it into 10-km strips of territory, and assume no overglow across sites (if the square sites are large enough), I posit

$$\hat{u}_{s,d} = u_{s,d} + \sum_{j=1}^{D} \rho_j \hat{u}_{s,d\pm j} + \varepsilon_{s,d} \tag{2.3}$$

where $u_{s,d}$ is the amount of light generated in site s and strip d, and $\hat{u}_{s,d}$ is the amount of light observed in site s and strip d. All strips of land in a site are ordered, so that strip 0 is adjacent to strips 1 and -1, which in turn are adjacent to strips 2 and -2 respectively, and so on.

The variable $u_{s,d}$ is unobserved, which prevents me from estimating equation 2.3 without additional assumptions. A straightforward approach would be to look for regions in the globe where I can construct strips such that $u_{s,d} = 0$ for some strips d. One such approach could be to use overglow into oceans; however, there is reason to believe that the reflective properties of light over oceans (and hence, the parameters of the overglow relationship in equation 2.3) are likely to be different from the reflective properties of light over land. Instead, I look at the edges between land subject to some kind of economic development and economically unexploited wasteland. Figure 2.6 shows the wasteland areas of the globe as defined by CIESIN in black: they are mostly deserts (Sahara, Arabian peninsula, Kalahari, Central Asia, the Australian Outback), rain forests (Amazon, Congo basin), mountains (United States, Chile) and tundra (Siberia and the Canadian north).

I break up the world map into 1×1 degree squares and extract those that contain boundaries between wasteland and non-wasteland areas. I then break up each square into 10-km wide strips parallel to the wasteland boundary and estimate equation 2.3 for wasteland strips in each square, assuming that for these strips, $u_{s,d} = 0$. Table 2.2 presents my results from assuming various orders of autoregression in the overglow relationship captured by equation 2.3. We see that the first-order autoregression is large and statistically significant, while subsequent orders are smaller in magnitude and only marginally statistically significant if at all. In particular, I fail to reject the null hypothesis that all autoregressive coefficients beyond the first are joinly zero unless I use a 10% significance level, in which case I marginally reject this null hypothesis in one of the specifications. Hence, a reasonable assumption for the overglow process is that it is first-order autoregressive, with the autoregressive coefficient being about 0.23.

I implement my overglow correction by inverting equation 2.3 to recover the amounts of light generated at the strips of land near the borders. Specifically, I obtain

$$u_{b,d} = \max\left(\hat{u}_{b,d} - 0.23 * (\hat{u}_{b,d-1} + \hat{u}_{b,d+1}), M\right) \tag{2.4}$$

where M is the recorded minimum value of light density, the censoring done to prevent negative values of light density.

2.4.4 Baseline Results

I now present formal analysis to document economic discontinuities at borders. I first run regressions at each border piece in each year of the form

$$y_{i,b,t,d} = \tilde{y}_{i,b,t} + \tilde{\delta}_{i,b,t} d + \eta_{i,b,t,d}, \text{ weighted by } 1\,(d \leq h)$$

where $y_{i,b,y,d}$ is the value of the dependent variable (lights per capita, or growth of lights per capita) in country i, year t, border piece b and distance d away from border b. The parameters to be estimated are $\tilde{y}_{i,b,t}$, the value of y at the border in country i, and $\tilde{\delta}_{i,b,t}$, the slope of y at the border. I choose the bandwidth h to assign greater weight to observations close to the border using a cross-validation procedure that selects the bandwidth to maximize the accuracy of predictions 10 km away from the border for each regression I run.

Having obtained estimates of log lights per capita and its slope at borders, I run regressions of the form

$$\tilde{y}_{i,b,t} = \alpha^w_{b,t} + \beta^w u^w_{i,b,t} + \varepsilon^w_{i,b,t} \tag{2.5}$$

$$\tilde{y}_{i,b,t} = \alpha^w_{b,t} + \gamma^w y^w_{i,t} + \varepsilon^w_{i,b,t} \tag{2.6}$$

where $\tilde{y}_{i,b,t}$ is the local linear estimate of log lights per capita at border piece b in country i and year t, $y^w_{i,t}$ is a measure of log output per capita (using method w, lights or national accounts) for country i as a whole, $u^w_{i,b,t}$ is an indicator that country i has the larger log output per capita (measured by w) of the two countries at border piece b, $\alpha^w_{b,t}$ is a border piece-year fixed effect, and $\varepsilon^w_{i,b,t}$ is the error term. The parameters of interest to be estimated are β, the average percentage rise in output per capita as one crosses a border from a poorer to a richer country, and γ, the elasticity of the ratio of output per capita at the border to the ratio of output per capita of the bordering nations. I weigh all observations by the population in a 70-km neighborhood of the border, and I cluster all standard errors by border (not border piece). Finally, I augment the variance of the regression by the first-step variances of the dependent variables from the local linear estimation according to the formula:

$$\bar{V} = \frac{N}{N-K}\left(\hat{X}'W\hat{X}\right)^{-1}\left(\hat{X}'W\left(\operatorname{diag}\left(\hat{\varepsilon}^2\right)+\bar{V}\right)W'\hat{X}\right)\left(\hat{X}'W\hat{X}\right)^{-1}$$

where N is the number of observations, K is the number of regressors (including fixed effects), $\operatorname{diag}(\hat{\varepsilon}^2)$ is a diagonal matrix of the squared residuals, W is a weight matrix, $\hat{X}$ is the matrix of regressors including the fixed effects, and $\bar{V}$ is a diagonal matrix of the first-step local linear estimation variances.

The first four columns of table 2.3 presents estimates of β and γ for different choices of the national output series as well as robustness and placebo checks. The first row shows the baseline estimates, in which I have corrected the data for overglow before computing the local linear estimates. We see that when national output is measured by log lights, light density jumps on average by 0.58 log points or nearly 80%, upon crossing from a poorer country into a richer one. About 65% of the difference in log output per capita between two bordering countries persists up to their joint border. Both of these findings are significant at 5%. If we look at discontinuities in light density at borders when national output is measured by GDP per capita, the results are slightly muted: light density jumps on average by only 0.4 log points (49%, and this is significant only at 10%), but 63% of the difference in log output per capita between bordering countries persists to the border. The average bandwidth used to obtain these estimates is large, about 51 km (I restrict my analysis to a 70-km neighborhood of the border). To assess the sensitivity of my estimates to bandwidth choice, I provide results with a bandwidth of 30 km for all countries. The coefficients β in the specifications with indicator variables (2.5) decline and lose significance, but the coefficients γ in the elasticity specifications (2.6) remain significant at 5%, though their magnitude is somewhat smaller. While the sensitivity of my result for indicator variables to the bandwidth is concerning, it is intuitive that the elasticity specifications are more flexible in accounting for the fact that discontinuities at different borders are of different size than are the indicator variable specifications. Moreover, any residual overglow after the correction is a much larger problem for a smaller bandwidth than for a larger one.[9]

A natural falsification exercise for my results is to re-run my regressions using fake borders, for which I should not expect a discontinuity. I perform two such exercises: one in which I draw the fake borders at a distance of 30 km from the real borders into the interior of the richer country, and one in which I draw the fake borders at a distance of 30 km into the interior of the poorer country. All the discontinuity estimates at the fake borders are less than half the size of the baseline estimates and statistically insignificant (although some of the fake indicator estimates are close in magnitude to the indicator estimates for the 30-km bandwidth), which gives reassurance to the hypothesis that there is something meaningful about national borders that creates discontinuities at them.

In the remaining rows of Table 2.3 I demonstrate both the importance and the plausibility of the overglow correction. A telltale feature of overglow should be that the slope of light density on the poorer side of the border should be negative (because light density is rising towards the border through contamination from the richer country) and that the slope of light density on the richer side of the border is positive (because light density is falling towards the border as there is no reinforcement of light from the poorer country). Hence, the slope of the light density from the local linear estimation on a given side of that border should be negatively associated with the output on that side. In the second row of Table 2.3 I run the regressions (2.5) and (2.6) that produce my baseline estimates, but I use the local linear slope rather than the intercept as the dependent variable (standardizing it for ease of interpretation). For my baseline estimates, the correlation between the local linear slope and output at each side of the border is insignificantly different from zero, and low in absolute value, though positive, suggesting some residual overglow. In rows 6-9, I present discontinuity estimates for log light density and its slope without the overglow correction, and with a cruder overglow correction in which I simply omit the last 10 km before the border from my analysis. The last correction is not preferred, because along with the overglow, it ignores any convergence in the level and growth rate of economic activity that might be going on over these last 10 km. We see that without an overglow correction, the discontinuities in log light density across borders are significant only at 10%, and substantially lower in magnitude than in my baseline results. However, the local linear slopes are radically higher on richer sides of borders – by as much as 0.3 standard deviations – which is consistent with substantial overglow. Correcting crudely for overglow by omitting the last 10 km yields discontinuity estimates much closer to the baseline, although the local linear slopes are still statistically significantly higher on richer sides of borders (though not by as much as without the overglow correction). Hence, we see that the data is consistent with the hypothesis that overglow is muting discontinuities in economic activity at

[9] In results not reported, I estimate equations 2.5 and 2.6 in which the weights are divided by the estimation variance from the discontinuity step for each observation in order to obtain feasible GLS estimates. For borders with very low estimation variance (e.g. borders with zero light density), I censor the reciprocal of the estimation variance at a fixed value, and about 27% of the observations (country-border piece-year) are so censored. The point estimates are similar to the ones reported (in fact, slightly larger) and the t-statistics of the point estimates increase, but the FGLS results cannot be distinguished from the baseline results by the Hausman test.

borders, and that correcting for overglow in various ways yields estimates of similar magnitude.

Since national-level variables are expected to affect GDP per capita in some manner, and since national borders change very infrequently, it may not be very surprising that there is a discontinuity across borders in GDP per capita produced by a long-term and consistent operation of national-level variables. What is more surprising to find is that 20-year growth rates in GDP show a similar discontinuity. Given that differentials in growth rates between countries are much less persistent than are differentials in GDP per capita, such a result suggests that national variables affect output rapidly and profoundly enough for changes to be noticeable over short periods of time.The last four columns of Table 2.3 present similar discontinuity estimates for average annualized 20-year growth rates. Here, the parameter β measures the average percentage point difference in annualized 20-year growth rates between higher-growing and lower-growing countries at borders, and the parameter γ measures the fraction of the percentage point differences in nationwide growth rates of two bordering countries that persists to the border. We see that discontinuities in economic growth are, if anything, even stronger than discontinuities in the level of economic activity. Crossing a border from a lower-growing into a higher-growing country is associated with a 2-3 percentage point rise in the average annualized growth rate. The differential between the nationwide growth rates of two bordering countries, on average, persists up to the border almost completely, or actually increases near the border. Using a smaller bandwidth, if anything, strengthens these results, and the falsification exercises with the fake borders show no discontinuities away from the true national borders. Overglow appears to be a smaller problem for measuring discontinuities in growth rates than in levels: the crude overglow correction of dropping the last 10 km before the border produces insignificant differentials in the slopes of growth rates on higher- and lower-growing sides of borders just as the more comprehensive overglow correction discussed in the text.

2.4.5 Results by Continent

It is important to understand which borders contribute most to my baseline finding. Table 2.4 presents estimates of discontinuities in log light density and in the growth rate of light density at borders between OECD countries, between post-Communist countries and other European countries, in Asia, Africa and in the Americas.[10] It is apparent that the borders with the strongest discontinuity estimates are Asian borders and the borders of post-Communist European countries, followed by borders in the Americas and among OECD countries. One noticeable fact is that border discontinuities are much weaker in Africa than everywhere else. One reason why Africa may have much smaller border discontinuities is that as a result of Africa's colonial experience, borders of African nations tend to be relatively underdeveloped hinterlands, with most political and economic activity concentrated around capital cities. Therefore, African nations having heterogeneous government activity in their heartland may have similar government activity (i.e. none) at their borders, and therefore, they may fail to exhibit border discontinuities.

2.4.6 Robustness Check to Geographic Controls

One potential problem with the results may be if national borders tend to be drawn at discontinuous changes in geographic variables, such as altitude, slope of the terrain (Nunn and Puga's

[10]The Americas are treated as a single continent because they are contiguous, and the number of borders in the Americas is comparable to that in Europe, Asia or Africa. Since the latter three continents form a continuous landmass, I define Russia, Turkey, Egypt and Israel to be part of two continents at once. I count the borders of these countries to belong to the continent of the bordering country: e.g. Russia's border with Estonia is a European border, but Russia's border with China is an Asian border. I count the Egypt-Israel border to be an Asian border.

[2009] ruggedness), and climatic zones (because differences in altitude may lead to differences in soil type, temperature and precipitation). To address that problem, I calculate average values of altitude, slope, and statistics of temperature and precipitation (means, maximums, minimums, variability measures) on either side of every border, obtained from WorldClimate at 30 arcsecond resolution. I also calculate, on either side of every border, the fraction of land belonging to each of the 14 climatic zones defined by the International Geosphere Biosphere Programme, by using the Global Land Cover Characteristics Data available from USGS, also at 30 arcsecond resolution [11]. I include these covariates in the regressions (2.5) and (2.6) as control variables on the right-hand side. I present estimates of the main coefficients of interest from both the levels and growth regressions in the second row of Table 2.5. It is clear that including the covariates does not change the results, and, if anything, strengthens them slightly. I also rerun the specifications given by equation (2.5) replacing the dependent variable by each of the 35 geographic variables used. Out of 140 possible tests, exactly 8 reject at 5%, which is approximately the number that one would expect if none of the geographic variables were discontinuous. I present the local linear regressions for some geographic variables thought to be potentially important for economic growth in Table 2.6: population density, altitude, ruggedness, mean temperature, temperature standard deviation, mean precipitation and fraction of land that is cropland. All the variables are scaled to have mean zero and standard deviation equal to unity. We see that richer sides of borders may be more rugged than poorer sides are, which is counterintuitive because ruggedness is typically associated with lower economic activity (Nunn and Puga 2009). They also have a lower degree of seasonal temperature variation, which is more alarming as a more moderate climate may improve economic activity, but as the magnitudes of the coefficients show, this tendency is very low even though statistically significant. All the other important measures show no tendency to be different on richer sides of borders from poorer sides of borders. There is also no tendency of these measures to be different on higher-growing sides of borders and lower-growing sides of borders.

2.4.7 Local Variation in Public Goods

I now consider potential explanations for discontinuities in economic activity across borders. First, I check whether public goods, for which extensive geographic information exists, are discontinuous across borders. If public goods are more extensively provided on richer sides of borders than on poorer sides of borders, it is conceivable that the output and growth differentials between the two sides are explained by the local effects of public goods. An example of local public goods driving income differentials is found by Dell (2010), who documents that road density falls discontinuously as one crosses a border into a Peruvian region in which forced labor was practiced during colonial times, and provides anecdotal evidence that road quality in that region is critical for access to markets.

I obtain spatial data on roads, railroads and utility lines (power and telephone lines) from

[11]The climatic variables are: (1) Annual Mean Temperature, (2) Mean Diurnal Range, (3) Isothermality, (4) Temperature Seasonality, (5) Max Temperature of Warmest Month, (6) Min Temperature of Coldest Month, (7) Temperature Annual Range, (8) Mean Temperature of Wettest Quarter, (9) Mean Temperature of Driest Quarter, (10) Mean Temperature of Warmest Quarter, (11) Mean Temperature of Coldest Quarter, (12) Annual Precipitation, (13) Precipitation of Wettest Month, (14) Precipitation of Driest Month, (15) Precipitation Seasonality, (16) Precipitation of Wettest Quarter, (17) Precipitation of Driest Quarter, (18) Precipitation of Warmest Quarter, and (19) Precipitation of Coldest Quarter. The land cover categories are: (1) Evergreen Needleleaf Forest, (2) Evergreen Broadleaf Forest, (3) Deciduous Needleleaf Forest, (4) Deciduous Broadleaf Forest, (5) Mixed Forest, (6) Closed Shrublands, (7) Open Shrublands, (8) Woody Savannas, (9) Savannas, (10) Grasslands, (11) Permanent Wetlands, (12) Croplands, (13) Urban and Built-Up, (14) Cropland/Natural Vegetation Mosaic.

the Digital Chart of the World. In results not reported, I confirm that the national aggregates of roads and railroads match up well with national-level statistics presented by the World Bank, and that there is a strong positive association between the log total public goods in a fixed neighborhood of a border and the amount of lights per capita in that neighborhood. In Table 2.7, I obtain estimates of discontinuities in the densities of these public goods variables per capita across national borders, which I construct using the two-step local linear procedure described in Section 2.4. I normalize the dependent variable to have mean zero and unit variance in all specifications. We see that road density is not systematically higher on the sides of borders corresponding to richer- or higher-growing countries (the estimates of the difference are positive, as might be expected if higher road density is associated with higher economic growth or development, but they are quantitatively small and statistically insignificant). An important component of public goods provision not captured by the road density maps is road quality. The Digital Chart of the World classifies roads as "primary or secondary roads," "divided highways" and "paths or trails." Presumably, trails are of lower quality than are the other two types of roads, so the fraction of roads that are not trails should be a reasonable measure of road quality. However, as we see in column 2, the fraction of roads that are not trails does not vary discontinuously across borders. Column 3 of Table 2.7 presents the results for railroads. The estimates are larger, but not significant (except the estimate for differences in GDP growth rates, which is significant at 10%). Finally, column 4 presents the same analysis with the dependent variable being log total length of utility lines (most frequently these are power lines, but also included are telephone lines and pipelines). The results are even more striking because they show that poorer sides of borders have more utility lines than richer sides of borders do, the result being statistically significant. Lower-growth sides of borders also have more utility lines than do higher-growth sides of borders, though this result is not statistically significant. This finding is particularly surprising given that the electricity used to generate the lights almost certainly is going through the power lines being measured. A reconciliation of this finding with the baseline result may be that the power lines are used with different intensity on different sides of borders, with more (and more energy-intensive) houses and factories using the power lines on richer sides of borders than on poorer sides.

It is instructive to look at maps of road density for the regions in which we observed significant discontinuities in the level and growth rate of light density per capita. In contrast, these regions appear to have continuous road density across borders. Superimposed on a road map of the Korean peninsula, the North Korea-South Korea border is imperceptible. In Eastern Europe, road networks rarefy in the sparsely populated Carpathian mountain region that straddles Ukraine and Romania, but have the same density on either side of the Ukraine-Poland or the Romania-Moldova borders. Hence, it is plausible that road networks are continuous across borders more generally, and hence, that discontinuities in economic activity at borders are not accounted for by variation in road networks.

Finally, I estimate the baseline regression with levels of public goods being included as independent variables. Rows 2 and 4 Table 2.5 present the results for my baseline specification with the different varieties of public goods included as independent variables, either by themselves (in row 2) or together with geographical and climate controls (row 1). The coefficients on the country-level output variables are very similar in magnitude and significance to the baseline in Table 2.3. Therefore, variation in local public goods is unlikely to account for border discontinuities in economic activity. However, variation in the national level of public goods still may be important if the main benefits from public goods are global rather than local (e.g. a high national level of public goods enriches everyone sufficiently that places with low local public goods benefit from having richer trading partners).

2.5 Potential Mechanisms

I now turn to exploring the association between discontinuities in the growth rate of lights per capita at borders and national-level variables of the bordering countries that are commonly hypothesized to be determinants of economic growth. This analysis will not use any further sources of identification and will be much closer to a correlational study of determinants of growth. However, because areas close to national borders can be considered as having been "randomly assigned" vectors of determinants of growth, whereas countries as a whole obviously have developed these vectors endogenously over the course of history, an important source of endogeneity will be absent. If an exhaustive list of determinants of growth existed, the identification problem in trying to explain discontinuities in economic activity at the border would be solved. While such a list does not exist, accounting for its most important components is conceivable.

The past decade has been extremely fruitful in investigating how government activity may affect economic growth. Acemoglu et al. (2000, 2001) argued that the degree of property rights protection and the constraints faced by the country's executive have a strong causal impact on economic activity. La Porta et al. (1998) and, less formally, de Soto (2000) have suggested that contracting institutions such as the simplicity of procedures to enforce contracts and trade property improve economic growth. However, Acemoglu and Johnson (2005) argue that controlling for property rights institutions, contracting institutions do not matter for growth. Glaeser et al. (2004) and more recently La Porta et al. (2008) contend that an important determinant of economic growth is human capital, which in almost all countries is substantially affected by government policies. Algan and Cahuc (2010) find a causal effect of trust on growth, and a line of thought stretching back to Weber (1910) posits that cultural attitudes formed by religion affect economic growth. Since government policies tend to homogenize religion and culture within their borders, these variables may also be discontinuous across borders, and thus, may be potential channels through which border discontinuities in economic activity arise.

While the literature has identified institutions of various kinds to be a fundamental cause of economic growth, it has not yet fully explained the channels through which these institutions bring about growth. One strand of thought, dating back to the founding of economics as a science, has considered that security of property enables citizens to optimize their well-being by trading and producing in the market, bringing about prosperity without much further action from the government except to enforce the property rights system. In the words of Adam Smith (1776) : "Little else is required to carry a state to the highest degree of affluence from the lowest barbarism but peace, easy taxes, and a tolerable administration of justice; all the rest being brought about by the natural course of things." An alternative theory, however, may be that secure property rights and tight constraints on the executive enable citizens to control their government and ensure that it provides adequate amounts of high-quality public goods, the latter being necessary for development and long-run growth. While I am unaware of recent work postulating this theory, the literature on economic growth and political economy has established both theoretically and empirically that a causal channel of institutions on growth that runs through public goods is conceivable. Dell (2010) and Huillery (2009) document the importance of public goods for positive economic outcomes. Acemoglu (2005) presents a model in which "consensually strong states," in which the citizens are strong relative to the government, collect higher taxes and spend more on public goods (as the citizens desire) than states in which government is not checked by citizens.

The regression of interest is

$$\tilde{g}_{i,b}^{w} = \alpha_b + \gamma_g^w g_i^w + Z_i'\Delta + \varepsilon_{i,b}^w \tag{2.7}$$

where Z_i is a vector of country-level determinants of growth and Δ is a vector of their coefficients, which, together with γ_g^w, the correlation between differences in growth at the border and differences in national growth, form the variables of interest. All other variables are defined as in equation (2.6). I use the 20-year growth rate of Penn World Table 7.1 GDP as my independent variable for the growth rate.[12]

The determinants of growth that I consider are as follows: the measure of the rule of law obtained from the World Governance Indicators, which are maintained by the World Bank; fraction of roads paved (a public goods variable motivated by Gennaioli and Rainer [2007]) obtained from the World Bank's World Development Indicators (WDI) for 1990 or the closest year; the amount of time necessary to enforce a contract (a measure of contracting institutions), also from the WDI for 1990; an index of political freedom based on the Freedom House measure, obtained from Acemoglu, Johnson, Robinson and Yared (2008); average years of education in 1990, obtained from Barro and Lee (2010); and the proportion of respondents from the given country answering that "most people can be trusted" in the World Values Survey, from La Porta et al. (2008). The governance measures, public goods and contracting institutions variables span nearly the whole sample of countries, while other variables tend to be missing for many countries, which removes borders from analysis.

A key variable in the analysis will be the rule of law variable from WGI, which is defined as "the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence." This variable (as well as the other WGI indicators) has been constructed using an unobserved components model using expert evaluations and surveys of businesses, NGOs and other agencies. A list of the questions in the surveys that were used to construct this variable is available at

http://info.worldbank.org/governance/wgi/pdf/rl.pdf

The questions from these surveys used to construct the rule of law variable ask about security of property, the willingness of the government to honor its contractual obligations and follow its own laws, and the independence of the judiciary, as well as the ease of using the judicial system to enforce private contracts, which makes the rule of law variable a composite of a measure of private property protection and of a measure of contracting institutions. However, since most of the questions used do not concern the enforcement of private contracts, I consider that the rule of law variable is much closer to a measure of private property institutions than it is to contracting institutions. I use the rule of law variable in this analysis rather than other measures of property rights institutions such as protection against expropriation risk from Political Risk Services or the property rights variable from the Fraser Institute because the rule of law variable is available for many more countries than either of these two variables. In Appendix Tables 2.A1 and 2.A2, I estimate the specifications in (2.8) using modified versions of these measures (with missing values for one measure predicted using the other measure) and obtain very similar results.

Table 2.8 presents results from estimating equation (2.7) for several versions of the vector of determinants Z_i using the two-step local linear procedure described in Section 2.4. For comparability purposes, all covariates (except the World Bank growth rate) have been standardized before regression, so the unit of each covariate is a standard deviation. Column 1 reproduces the

[12]The results are not robust to using the nationwide 20-year growth rate of light density in place of the 20-year Penn World Tables growth rate. However, if the climate and local public goods controls are included, the results are qualitatively similar to the GDP-based results, although none of the coefficients except for the 20-year nationwide lights growth rate are significant.

baseline estimate of γ_g, the baseline association between border and national differences of growth, for comparability purposes. Column 2 includes the rule of law as a covariate. We see that the coefficient on the rule of law is positive and statistically significant, with a 1 standard deviation increase in the difference of bordering countries' expropriation protections being associated with an increase in the difference of their average annualized 20-year growth rates at the border of 2.5 percentage points. The coefficient on the difference in national growth rates is halved (to 0.44) and becomes only marginally statistically significant.[13] Column 3 leaves out the rule of law, but includes public goods (percent of roads paved). We see that public goods with a coefficient of about 1.8, hence a 1 standard deviation increase in the difference between two countries' fractions of roads paved increases the difference of their 20-year growth rates at the border by 1.8 percentage points. However, this coefficient is not statistically significant at conventional levels, and the coefficient on the national growth rate, γ_g, remains significant, with nearly the same magnitude as in the baseline specification (0.74). Column 4 includes both rule of law and public goods. We now see that the coefficient on the rule of law remains very close to the one from the second specification (2.31) and is significant at 1%, while the coefficient on public goods shrinks markedly to 0.67. The coefficient on the national growth rate shrinks to a marginally significant 0.44. Thus, controlling for the rule of law, public goods do not appear to matter for growth. Column 5 adds the climate and local public goods covariates at borders from Table 2.5. The estimate of the rule of law coefficient is unaffected, but the paved roads coefficient turns negative (though it remains insignificant) and the coefficient on the countrywide 20-year growth rate becomes slightly larger (0.54) and significant at 5%. In subsequent columns I retain the climate and local public goods controls to reduce the standard error of the regression; my results do not change if these controls are excluded. These results, especially column 4, show that of two ways that institutions that can matter for growth – creating a rule of law to protects private property in the market, and creating a consensual state to provide public goods – only the rule of law appears to matter for economic growth at national borders, in accordance with Adam Smith. Moreover, the rule of law is such an important variable that once differences in the respect for the rule of law are accounted for, differences in national growth rates are of limited use in predicting differences in border growth rates.

The additional columns of Table 2.8 present further checks of the importance of property rights protection compared with other determinants of growth. Column 6 compares private property institutions and contracting institutions by including the amount of time necessary to enforce a contract as a covariate; the coefficient is positive (counter to expectations) and statistically insignificant. The coefficient on the rule of law variable remains unchanged. Another hypothesis can be that property rights protection is proxying for political freedom. Column 7 replaces the contracting institutions variable with average value of the Freedom House Political Rights Index between 1990 and 1999. We see that, if anything, political freedom appears to decrease growth rates at the border once economic freedom is controlled for.[14] The coefficient on the rule of law variable rises by over 50% to 3.9. However, the Freedom House Political Rights Index is not available for some countries, which reduces my sample of borders by about one-third. In Column 10, I add the estimate of average years of education from Barro and Lee (2010) to the specification in

[13] The magnitudes of the two coefficients are not comparable because expropriation protection has been standardized while the national growth rate has not. I estimate this equation using beta coefficients, and find a beta coefficient of 0.1 for the national growth rate and a beta coefficient of 0.25 for expropriation protection. Therefore, the magnitude of the association between expropriation protection and border growth, conditional on border fixed effects, is about 2.5 times larger than the magnitude between the association between national growth and border growth when both are expressed in like units

[14] The coefficient on the Freedom House variable is negative and statistically significant at 5%, but is only significant at 10% if the climate and local public goods covariates are excluded.

Column 5. Like the Freedom House measure, the Barro-Lee education measure is not available for some borders, which reduces my sample. The coefficient on the rule of law variable rises somewhat compared to Column 5 (to 3.3), and the coefficient on the Barro-Lee variable is negative, large in magnitude, and statistically significant at 5% (−2.7). Finally, column 9 includes a variable for trust measured by the World Values Survey.[15] Unfortunately, data on trust from the WVS is available for only 80 countries, which severely restricts the sample. Therefore, I construct an augmented trust variable by predicting trust using region dummies for countries without trust data. Given that cultural variables tend to be similar within regions (see the map by Ingelhart and Welzel, http://www.worldvaluessurvey.org/), this approach appears to be reasonable. The coefficient on the rule of law remains very close to the baseline, and the coefficient on trust is negative and insignificant. In all the above specifications, the coefficient on the fraction of roads paved is statistically insignificant, and frequently negative. Hence, the rule of law appears to be resilient to controlling for other potential determinants of growth.

2.5.1 Border Permeability

An important concern for interpreting border discontinuities as estimates of the impact of government activity upon the economy is that national borders act as discontinuous increases in transaction costs. Even in a world with countries having identical institutions, public goods provision and other political structures, we could then still see discontinuities in economic activity at borders because local economic shocks would not transmit to neighboring countries but would remain contained in the country of origin. If borders were completely impermeable to trade, different market equilibria would prevail in each country, and shocks that affect one country's equilibrium would have no effect on its neighboring country, thus creating a discontinuity in prices, wages, quantities, and most likely economic activity and growth across the countries' common border.

I test whether the transaction costs channel is important by looking at whether the presence of trade mitigates discontinuities in economic activity at borders, and whether accounting for trade changes the ways in which the covariates discussed in this section affect discontinuities at borders. Since the transaction costs channel should manifest itself through differences in trade flows, if the amount of trade across the border (suitably normalized to account for obvious determinants of trade) does not affect whether or not border discontinuities are present, then it is unlikely that borders pose sufficient barriers to trade to explain the existence of border discontinuities. For the key independent variable of this part of the analysis, I obtain data on bilateral trade for all borders in 2000 from the IMF's Direction of Trade Statistics. As my trade variable, I use the log volume of trade, normalized by the product of the bordering countries' GDPs to take into account gravity effects. This trade variable is obviously at the border level, and hence is captured by border fixed effects, so I use it only to create interaction terms in regressions.

The regression of interest becomes

$$\tilde{y}_{i,b,t} = \alpha^{w}_{b,t} + \left(\gamma^{w} + t^{w}_{\beta} \times T_b\right) y^{w}_{i,t} + Z_i' \left(\Delta + \tau_Z \times T_b\right) + \varepsilon^{w}_{i,b,t} \tag{2.8}$$

where T_b is the trade measure, and $t^{d,w}_{g}$ and τ are coefficients on the interactions of this trade measure with the national growth measure and with the covariates, respectively. All other variables are as in equations (2.6) and (2.7). For ease of interpretation, I standardize T_b to have mean zero

[15]The exact text of the relevant WVS question is: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" The two responses are "Most people can be trusted," and "Can't be too careful."

and variance 1.

Table 2.9 presents some key specifications from the previous sections estimated using the two-step local linear regression methodology in Section 2.4. The first four columns present results for reestimating the baseline equation (2.6) with trade interactions. The coefficient on national output per capita or national growth in output per capita decreases, and the interaction with trade enters negatively for the regressions of log output per capita (the coefficient being insignificant if lights data is used to construct the output measure and marginally significant at -0.23 if GDP data is used), which suggests that some of the discontinuities in output per capita might be explained by trade. However, the trade interaction coefficients are statistically insignificant, small and positive for the growth rate regressions, which contradicts the hypothesis that trade effaces border discontinuities, and hence, that border discontinuities can be explained by trade. The subsequent four columns replicate columns 2, 3, and 4 of Table 2.8, the analysis of the rule of law and public goods. We see that nothing important changes: the magnitudes of the interaction coefficients are tiny, and the stylized facts of the rule of law explaining border discontinuities and of public goods not explaining them as well remain.

In theory, the volume of trade should be a sufficient statistic for economic interaction between two countries, regardless of the source of the barriers to such interactions. However, trade flows are not a good proxy for economic interactions between two countries because of the difficulty of valuing trade in services, or because some trade might be informal and not recorded in official statistics. Appendix Tables 2.A3 through 2.A5 therefore provide estimates of equation 2.8 in which the trade measure T_b is constructed using average tariffs between two countries (obtained from the World Bank), migration between the two countries normalized by their populations (again, from the World Bank) and genetic distance between the two countries (from Spolaore and Wacziarg 2009). Note that the tariff barriers and genetic distance measures should be negatively correlated with border permeability, so positive rather than negative interaction coefficients suggest that higher border permeability decreases discontinuities at borders. We observe that discontinuities in growth rates, and their relationship to differences in the rule of law and public goods provision on different sides of borders, are not affected by accounting for differential border permeability (in fact, for the trade barriers measure, lower tariffs at borders appear to increase rather than decrease the magnitude of the discontinuities). The border permeability effects on discontinuities in levels are small (or the wrong sign) for some measures, but larger for others. One potential explanation may be that the economic channel for creating discontinuities at borders operates over a longer time horizon than does the political channel, since economic shocks may propagate slowly over a country, while changes in incentives brought about by changes in political institutions may be immediate. Hence, border permeability may not affect discontinuities in economic growth but might contribute to discontinuities in levels of economic development.

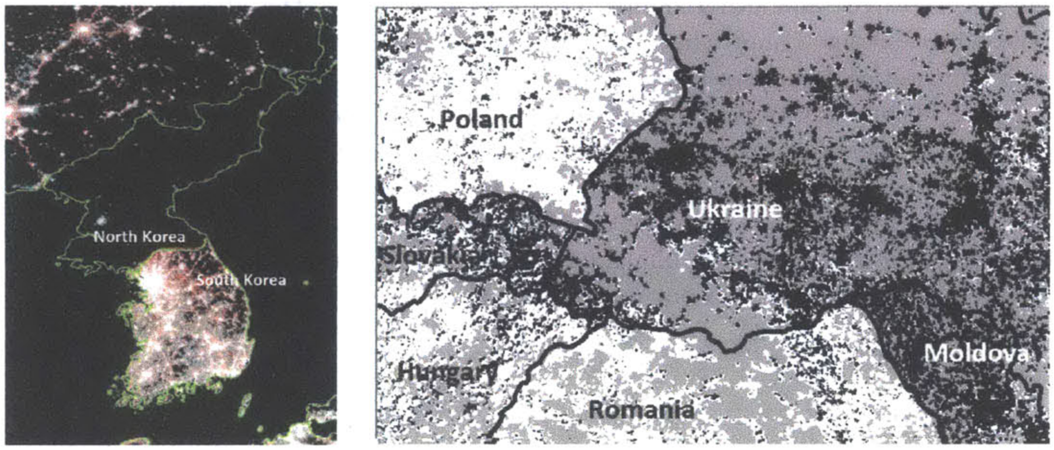
2.6 Conclusion

This paper uses satellite data on nighttime lights to find large and statistically significant discontinuities in economic activity at national borders. More surprising than the discontinuities in the levels of economic activity is the finding that there are equally large and significant discontinuities in growth rates of output over a 20-year period. Furthermore, I derive the properties of local polynomial estimators when the data generating process remains bounded but increases in resolution asymptotically, and use the derived variance formulae to compute local polynomial estimates of border discontinuities. In addition, I propose a novel procedure for removing overglow from nighttime lights data in a way that does not respect national borders and implement it to improve the performance of local linear estimation.

I interpret the estimated discontinuities in economic activity at the border as measures of the absolute effect of government activity (both short-run through policy and long-run through institutions) on economic activity. An alternative explanation for border discontinuities could be that they arise because borders discretely raise transaction costs, thus creating discontinuities between prices, wages and other nonpolitical variables across the border. Using data on trade flows between countries, I show that such an explanation is implausible because the magnitudes of border discontinuities do not seem to respond to differences in the flow of trade across the borders. I present a correlational analysis of potential determinants of border discontinuities, and find that the discontinuities in economic growth across borders can be explained by differences in the rule of law. In particular, once the rule of law is accounted for, differences in national growth no longer have a statistically significant association with differences in growth at the border, and other potential determinants of growth such as public goods, education, contracting institutions, political freedom and interpresonal trust either do not matter or do not account for the impact of the rule of law on growth.

2.7 Figures

Fig. 2.1 (2.1)

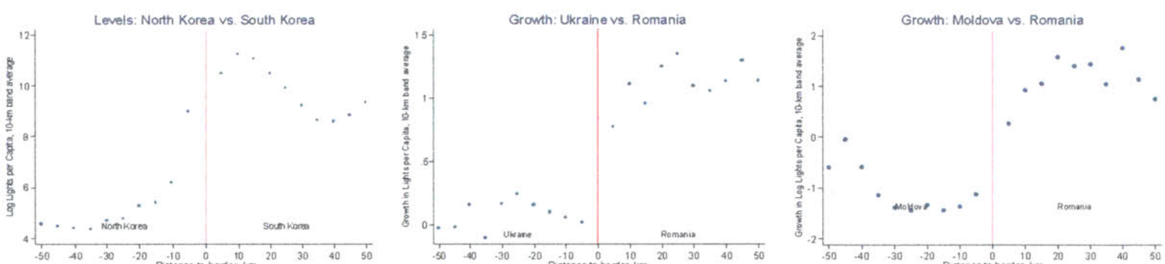


Korea, 2000 E. Europe, 1992-2000

(1) (2)

In (2), white areas denote lights increase and black areas denote light decrease

Fig. 2.2 (2.2)



Graphs of levels and growth discontinuities in lights.

Fig. 2.3 (2.3)

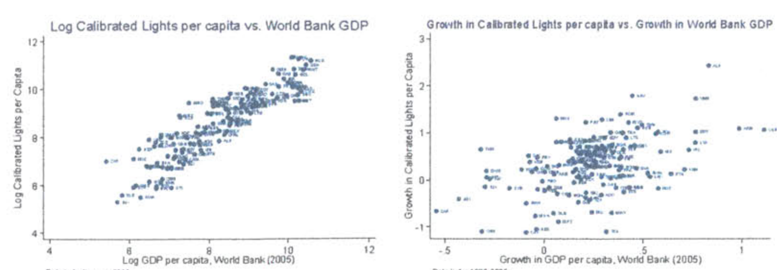
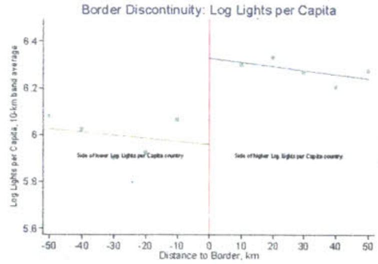
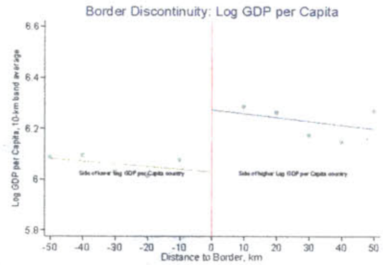


Fig. 2.4

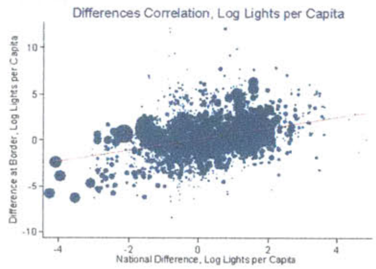
(2.4)



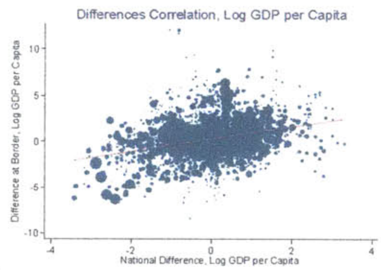
(1)



(2)



(3)

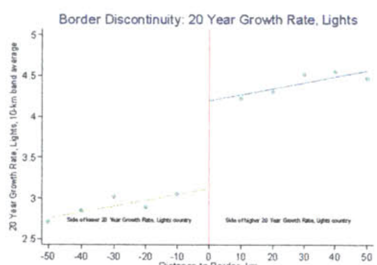


(4)

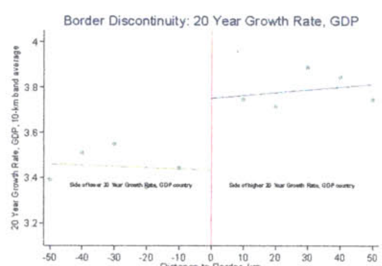
Discontinuity and Correlation Plots, Log Lights per Capita at Borders

Fig. 2.5

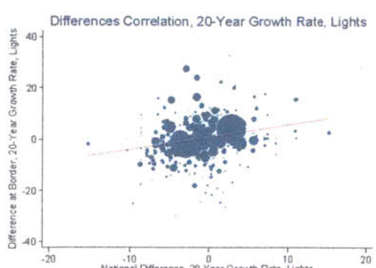
(2.5)



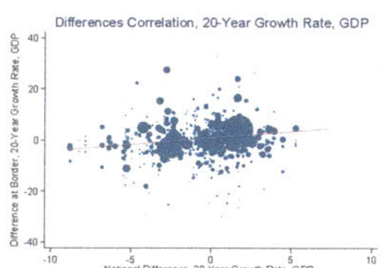
(1)



(2)



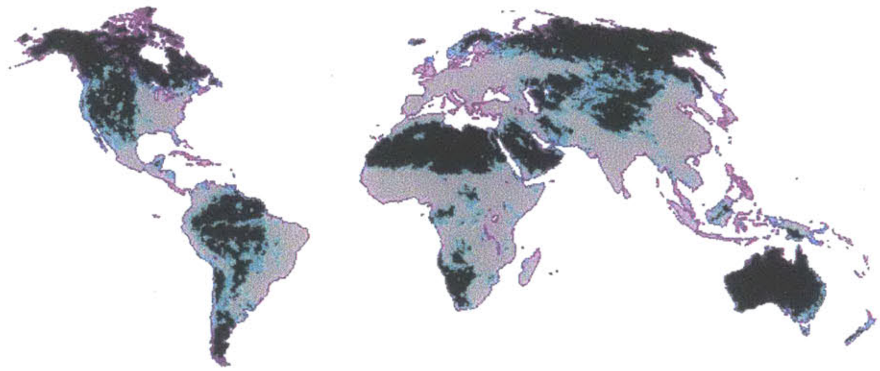
(3)



(4)

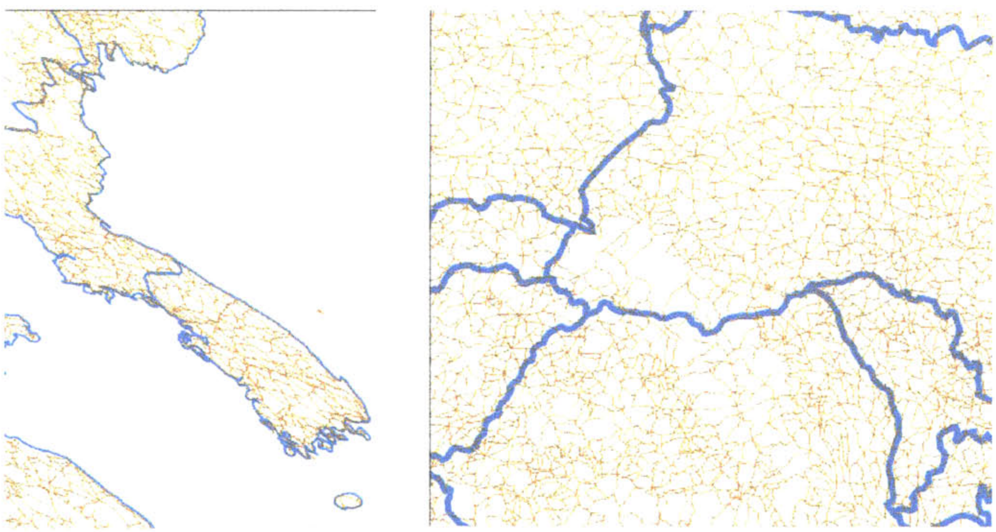
Discontinuity and Correlation Plots, Growth of Lights per Capita at Borders

Fig. 2.6 (2.6)



Areas denoted as wasteland by CIESIN shown in black.

Fig. 2.7 (2.7)



Road density on the Korean Peninsula and in Eastern Europe

2.8 Tables

Table 2.1

Descriptive Statistics				
Panel 1: Lights on the Border and Nationwide				
	Mean, Overall	Mean, Low Ntwide Side	Mean, High Ntwide Side	Number of Borders
Log Lights per Capita, Border	4.73	4.52	4.93	270
	(1.74)	(1.78)	(1.67)	
Log Lights per Capita Nationwide	8.00	7.58	8.41	270
	(1.40)	(1.47)	(1.19)	
Growth in Lights per Capita, Border	2.70	1.37	4.13	270
	(4.54)	(4.09)	(4.56)	
Growth in Lights per Capita, Nationwide	3.38	1.94	4.92	270
	(3.28)	(2.84)	(3.02)	
Log Lights per Capita, Border	4.73	4.59	4.86	270
	(1.74)	(1.76)	(1.70)	
Log GDP per Capita, WB	7.87	7.60	8.15	270
	(1.21)	(1.23)	(1.14)	
Growth in Lights per Capita, Border	2.70	1.95	3.47	270
	(4.54)	(4.18)	(4.76)	
Growth in GDP per Capita, WB	2.48	1.57	3.43	270
	(1.91)	(1.70)	(1.62)	
	Panel 2: Covariates			
	Mean, Overall	Mean, Low Ntwide Growth GDP Side	Mean, High Ntwide Growth GDP Side	Number of Borders
Log Roads in 30-km Border Neighborhood	3.02	2.96	3.09	270
	(1.67)	(1.74)	(1.60)	
Log Population	16.06	15.99	16.14	270
	(2.05)	(2.11)	(1.97)	
Rule of Law, WB	-.34	-.49	-.19	266
	(.91)	(.90)	(.89)	
Percent of Roads Paved, WDI	-.26	-.30	-.22	264
	(.90)	(1.00)	(.77)	
Average years of Education, BL	-.57	-.63	-.51	186
	(.88)	(.85)	(.91)	
Trust, WVS	.19	.03	.32	98
	(.83)	(.76)	(.86)	

(2.1)

Descriptive Statistics. Standard deviations are in parentheses. There are two observations per border piece: one for the poorer (or lower-growing) side, and one for the richer (or higher-growing) side. Data for lights at border and their growth rate are for 70-km neighborhoods around the border. Data for roads in 70-km neighborhood of border from Digital Chart of the World. Data for rule of law and fraction of roads paved from the World Bank and WGI. Data for education from Barro-Lee (2010). Data for trust from WVS.

Table 2.2

Overglow into Wasteland *Dep. Var. is Light Density in Wasteland*			
	(1)	(2)	(3)
Light Density, 10 km Away	.23***	.29***	.35***
	(.02)	(.05)	(.05)
Light Density, 20 km Away		-.03*	-.06*
		(.02)	(.03)
Light Density, 30 km Away			.02
			(.01)
No. Observations	2949	1996	1123
No. Squares	888	768	631
R^2	.55	.59	.63
P-value higher lags are 0		.1	.17

(2.2)

Overglow Correction. Each observation corresponds to a 1-degree grid square on a world map, split up into 10-km wide bands that are parallel to the frontier between wasteland and non-wasteland in that square. Wasteland areas are defined according to CIESIN as shown in Figure 2.6.

Table 2.3

Local Linear Estimates of Border Discontinuities								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	Log Lights p/c	Log Lights p/c	Log Lights p/c	Log Lights p/c	Growth, Lights p/c	Growth, Lights p/c	Growth, Lights p/c	Growth, Lights p/c
Indep. Var.	Dummy Log Lights p/c	Log Lights p/c	Dummy Log GDP p/c	Log GDP p/c	Dummy Growth, Lights p/c	Growth, Lights p/c	Dummy Growth, GDP p/c	Growth, GDP p/c
Baseline Estimates (H=51)	.58**	.65***	.40*	.63**	3.56***	1.29***	2.63**	.88***
	(.23)	(.23)	(.23)	(.29)	(.98)	(.30)	(1.07)	(.34)
Baseline Slope	.04	.04	.05	.10	-.20	-.07	-.12	-.03
	(.08)	(.08)	(.08)	(.11)	(.31)	(.09)	(.31)	(.10)
Estimates with BW=30 km	.28	.46**	.14	.56**	4.10***	1.49***	2.39**	.97***
	(.20)	(.22)	(.18)	(.28)	(.92)	(.28)	(1.02)	(.31)
Placebo Estimates at -30 km	.16	.05	.14	-.09	-.15	-.09	.37	.06
	(.25)	(.24)	(.25)	(.28)	(.70)	(.22)	(.67)	(.24)
Placebo Estimates at +30 km	-.02	-.01	-.04	-.07	-.12	.15	.40	.12
	(.19)	(.15)	(.19)	(.19)	(1.13)	(.33)	(1.09)	(.45)
Baseline Estimates, No Correction	.25*	.23*	.12	.23	1.11*	.46***	1.26**	.44**
	(.13)	(.13)	(.12)	(.14)	(.60)	(.15)	(.57)	(.20)
Baseline Slope, No Correction	.31***	.37***	.33***	.45***	.63***	.21***	.38*	.16*
	(.08)	(.08)	(.08)	(.10)	(.20)	(.06)	(.21)	(.08)
Baseline Estimates, 10 km off	.53***	.58***	.39***	.61***	2.41***	.93***	2.53***	.89***
	(.15)	(.16)	(.14)	(.17)	(.73)	(.21)	(.68)	(.28)
Baseline Slope, 10 km off	.18**	.19**	.20**	.26**	.00	-.01	-.20	-.01
	(.09)	(.09)	(.09)	(.11)	(.23)	(.08)	(.23)	(.11)
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Observations	6760	6760	6760	6760	1352	1352	1352	1352
No. Borders	270	270	270	270	270	270	270	270
R^2 of first row	.02	.03	.02	.03	.00	.00	.00	.00

(2.3)

Data on lights and population available from the NOAA and CIESIN, respectively, for 1990 (1992 for lights), 1995, 2000, 2005 and 2010. Observation unit is a country-border piece-year. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

Table 2.4

Border Discontinuities by Continent					
	(1)	(2)	(3)	(4)	(5)
Continent	OECD	Post-Soviet	Asia	Africa	America
Indicator, Log Lights per Capita	.30	.67***	.84**	.06	.75*
	(.22)	(.23)	(.42)	(.29)	(.42)
Log Lights per Capita	1.24*	1.30***	.80***	-.03	.70**
	(.66)	(.37)	(.31)	(.28)	(.35)
Indicator, Log GDP per Capita	.09	.34	.84**	-.30	.30
	(.26)	(.27)	(.38)	(.27)	(.34)
Log GDP per Capita	.67	.66*	1.14***	-.25	.66*
	(1.07)	(.38)	(.42)	(.27)	(.40)
Indicator, Growth Lights per Capita	.89*	10.44***	5.31***	-.25	2.16
	(.48)	(2.92)	(1.33)	(1.36)	(1.40)
Growth Lights per Capita	.93***	2.54***	1.66***	.19	1.93*
	(.32)	(.56)	(.45)	(.27)	(1.08)
Indicator, Growth GDP per Capita	-.61	6.53*	3.79**	1.11	-.18
	(.48)	(3.45)	(1.66)	(1.32)	(1.54)
Growth GDP per Capita	-.32	3.22**	1.16**	.14	1.42
	(1.13)	(1.62)	(.52)	(.33)	(1.10)
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
No. Observations	122	174	420	442	240
No. Borders	23	45	75	93	36
R^2 of first row	.00	.14	.08	.00	.04

(2.4)

Data on lights and population available from the NOAA and CIESIN, respectively, for 1990 (1992 for lights), 1995, 2000, 2005 and 2010. Observation unit is a country-border piece-year. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

Table 2.5

Border Discontinuities with Controls								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	Log Lights p/c	Log Lights p/c	Log Lights p/c	Log Lights p/c	Growth, Lights p/c	Growth, Lights p/c	Growth, Lights p/c	Growth, Lights p/c
Indep. Var.	Dummy Log Lights p/c	Log Lights p/c	Dummy Log GDP p/c	Log GDP p/c	Dummy Growth, Lights p/c	Growth, Lights p/c	Dummy Growth, GDP p/c	Growth, GDP p/c
Baseline Estimates	.58**	.65***	.40*	.63**	3.56***	1.29***	2.63**	.88***
	(.23)	(.23)	(.23)	(.29)	(.98)	(.30)	(1.07)	(.34)
Climate Controls	.70***	.90***	.47**	.69***	3.62***	1.28***	2.74***	.92***
	(.19)	(.20)	(.18)	(.25)	(.83)	(.26)	(.97)	(.28)
Local Public Goods Ctrls.	.60**	.68***	.40*	.62**	3.69***	1.30***	2.54***	.83***
	(.23)	(.25)	(.23)	(.30)	(.88)	(.27)	(.95)	(.31)
All Controls	.72***	.92***	.45***	.66***	3.68***	1.25***	2.53***	.84***
	(.18)	(.21)	(.17)	(.25)	(.75)	(.23)	(.90)	(.25)
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Observations	6760	6760	6760	6760	1352	1352	1352	1352
No. Borders	270	270	270	270	270	270	270	270
R^2 of first row	.04	.08	.02	.04	.12	.17	.06	.04

(2.5)

Data on lights and population available from the NOAA and CIESIN, respectively, for 1990 (1992 for lights), 1995, 2000, 2005 and 2010. Observation unit is a country-border piece-year. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. All control variables rescaled to have mean 0 and variance 1, and are described in the text. Each observation weighted by population in the respective buffer piece.

Table 2.6

Behavior of Geographical Covariates across Borders							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.	Log Population	Log Altitude	Log Slope	Mean Temperature	Log SD Temp.	Log Precipitation	Fraction Cropland
Indicator, Log Lights	-.01	.01	.12**	-.01	-.01***	.00	-.02
	(.04)	(.02)	(.05)	(.01)	(.00)	(.00)	(.03)
Indicator, Log GDP	-.02	.01	.09*	-.01*	-.00*	.00	-.03
	(.03)	(.02)	(.05)	(.00)	(.00)	(.00)	(.02)
Indicator, Growth Lights	-.01	.00	.06	-.00	-.00	-.00	-.03
	(.05)	(.03)	(.07)	(.01)	(.00)	(.01)	(.03)
Indicator, Growth GDP	-.05	.01	.09	-.01	-.00	.00	.00
	(.05)	(.03)	(.07)	(.01)	(.00)	(.01)	(.03)
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Observations	1352	1352	1352	1352	1352	1352	1352
No. Borders	270	270	270	270	270	270	270
R^2 of last row	.00	.00	.07	.02	.01	.00	.00

2.6

Data on lights and population available from NOAA and CIESIN. Data on altitude and slope obtained from SRTM. Data on climatic variables available from WorldClimate (Hijmans et al. 2005). All climate variables rescaled to have mean 0 and variance 1. Observation unit is a country-border piece-year. The last (10 km) observation on each side of the border is excluded to minimize overglow. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

Table 2.7

Behavior of Local Public Goods across Borders				
	(1)	(2)	(3)	(4)
Dep. Var.	Log Roads	Fraction Roads	Log Railroads	Log Utilities
Indicator, Log Lights	-.01	-.00	.12	-.64***
	(.06)	(.01)	(.14)	(.21)
Indicator, Log GDP	.03	-.01	.13	-.48**
	(.05)	(.01)	(.13)	(.21)
Indicator, Growth Lights	.07	-.01	.16	-.14
	(.08)	(.01)	(.16)	(.26)
Indicator, Growth GDP	.01	.00	.29*	-.25
	(.08)	(.01)	(.16)	(.25)
Border-Year Fixed Effects	Yes	Yes	Yes	Yes
No. Observations	1352	1352	1352	1352
No. Borders	270	270	270	270
R^2 of last row	.00	.00	.02	.00

2.7

Data on lights and population available from NOAA and CIESIN. Data on roads, railroads and road type available from the Digital Chart of the World for 2003. Observation unit is a country-border piece-year. All dependent variables rescaled to have mean 0 and variance 1. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

Table 2.8

Correlates of Border Discontinuities									
Dep. Var. is 20-Year Growth Rate of Light per Capita									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
20-Yr. Growth, GDP	.88***	.44*	.74***	.44*	.54**	.54	.35	.28	.52**
	(.34)	(.26)	(.26)	(.25)	(.25)	(.33)	(.37)	(.38)	(.26)
Rule of Law, WB		2.54***		2.23***	2.31***	2.34***	3.93***	3.32***	2.35***
		(.85)		(.69)	(.59)	(.60)	(1.12)	(.69)	(.63)
Frac. of Roads Paved, WB			1.80	.67	-.56	-.52	1.09	-.69	-.54
			(1.11)	(.91)	(.86)	(.91)	(.91)	(.90)	(.87)
Time to Enforce Contract						.47			
						(.79)			
Freedom House Score							-3.32**		
							(1.60)		
Schooling, Barro-Lee								-2.66**	
								(1.25)	
Predicted Trust, WVS									-.22
									(.62)
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clim. and Pub. Gds. Ctrls.	No	No	No	No	Yes	Yes	Yes	Yes	Yes
No. Observations	1352	1342	1336	1336	1336	1282	952	968	1336
No. Borders	270	266	264	264	264	254	180	184	264
R^2	.04	.10	.07	.10	.28	.29	.43	.38	.28

(2.8)

Data on lights and population available from NOAA and CIESIN. Data on determinants of growth described in the text. All covariates except 20-year growth normalized to have mean 0 and variance 1. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

Table 2.9

Permeability: Trade							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.	Log Lights	Log Lights	Growth, Lights	Growth, Lights	Growth, Lights	Growth, Lights	Growth, Lights
Nat. Output / Growth	.54**	.44*	1.33***	.87**	.46	.73**	.40
	(.23)	(.25)	(.32)	(.40)	(.34)	(.36)	(.35)
Nat. Output / Growth X Trade	-.15	-.28*	.10	-.00	.07	.13	.15
	(.12)	(.15)	(.17)	(.15)	(.15)	(.17)	(.15)
Rule of Law					2.62***		2.39***
					(.89)		(.76)
Rule of Law X Trade					-.36		-.24
					(.55)		(.48)
Fraction of Roads Paved						1.83	.53
						(1.16)	(1.00)
Fraction of Roads Paved X Trade						-.75	-.86
						(.80)	(.71)
Output Source	Lights	GDP	Lights	GDP	GDP	GDP	GDP
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Observations	6760	6760	1352	1352	1342	1336	1336
No. Borders	270	270	270	270	266	264	264
R^2	.09	.06	.17	.04	.10	.07	.11

(2.9)

Data on lights and population available from NOAA and CIESIN. Data on trade is from IMF's Direction of Trade Statistics for 2000. Trade volume normalized by product of bordering country GDPs and to have mean zero and unit variance. Observation unit is a country-border piece-year. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

2.9 Appendix I: Additional Tables

Table 2.A1

Correlates of Border Discontinuities: Expropriation Risk as Rule of Law Measure									
Dep. Var. is 20-Year Growth Rate of Light per Capita									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
20-Yr. Growth, GDP	.88***	.50*	.74***	.49*	.58**	.70**	.68**	.41	.59**
	(.34)	(.26)	(.26)	(.25)	(.24)	(.31)	(.34)	(.36)	(.25)
Avg. Protection ctr. Exp. AJR		2.54***		2.44***	2.55***	2.67***	3.32***	3.63***	2.55***
		(.89)		(.92)	(.82)	(.87)	(.87)	(.96)	(.81)
Fraction of Roads Paved, WB			1.80	.34	-.88	-.75	.96	-1.24	-.93
			(1.11)	(.87)	(.93)	(.95)	(.78)	(1.06)	(.94)
Time to Enforce Contract						.30			
						(.81)			
Freedom House Score							-1.84*		
							(1.11)		
Schooling, Barro-Lee								-1.97*	
								(1.06)	
Predicted Trust, WVS									.32
									(.57)
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climate and Public Goods Ctrls.	No	No	No	No	Yes	Yes	Yes	Yes	Yes
No. Observations	1352	1342	1336	1336	1336	1282	952	968	1336
No. Borders	270	266	264	264	264	254	180	184	264
R^2	.04	.14	.07	.14	.31	.32	.46	.43	.31

(2.A1)

Data on lights and population available from NOAA and CIESIN. Data on determinants of growth described in the text. All covariates except 20-year growth normalized to have mean 0 and variance 1. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

Table 2.A2

Correlates of Border Discontinuities: Alternative Rule of Law Measure *Dep. Var. is 20-Year Growth Rate of Light per Capita*									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
20-Yr. Growth, GDP	.88***	.69**	.74***	.64**	.71***	.76**	.75*	.59	.72***
	(.34)	(.27)	(.26)	(.25)	(.24)	(.32)	(.40)	(.37)	(.25)
Property Rights, Fraser Institute		2.00**		1.70**	1.78***	1.92***	1.91***	2.65***	1.78***
		(.82)		(.69)	(.56)	(.59)	(.65)	(.65)	(.56)
Fraction of Roads Paved, WB			1.80	.97	-.28	-.22	1.97*	-.31	-.30
			(1.11)	(.92)	(.88)	(.92)	(1.09)	(.93)	(.89)
Time to Enforce Contract						.55			
						(.80)			
Freedom House Score							-1.16		
							(1.25)		
Schooling, Barro-Lee								-1.99*	
								(1.16)	
Predicted Trust, WVS									.15
									(.57)
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climate and Public Goods Ctrls.	No	No	No	No	Yes	Yes	Yes	Yes	Yes
No. Observations	1352	1342	1336	1336	1336	1282	952	968	1336
No. Borders	270	266	264	264	264	254	180	184	264
R^2	.04	.09	.07	.10	.27	.28	.40	.38	.27

(2.A2)

Data on lights and population available from NOAA and CIESIN. Data on determinants of growth described in the text. All covariates except 20-year growth normalized to have mean 0 and variance 1. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

Table 2.A3

Permeability: Migration							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.	Log Lights	Log Lights	Growth, Lights	Growth, Lights	Growth, Lights	Growth, Lights	Growth, Lights
Nat. Output / Growth	.66***	.68**	1.40***	.99**	.59*	.87***	.59*
	(.24)	(.30)	(.30)	(.39)	(.30)	(.32)	(.30)
Nat. Output / Growth X Migration	.01	.11	.21	.15	.20	.17	.20
	(.11)	(.13)	(.13)	(.14)	(.16)	(.14)	(.15)
Rule of Law					2.86***		2.53***
					(.91)		(.77)
Rule of Law X Migration					.07		.07
					(.55)		(.55)
Fraction of Roads Paved						2.01*	.71
						(1.20)	(1.00)
Fraction of Roads Paved X Migration						.03	-.04
						(.70)	(.72)
Output Source	Lights	GDP	Lights	GDP	GDP	GDP	GDP
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Observations	6760	6760	1352	1352	1342	1336	1336
No. Borders	270	270	270	270	266	264	264
R^2	.08	.04	.18	.05	.11	.07	.11

(2.A3)

Data on lights and population available from NOAA and CIESIN, respectively, for 1990, 1995, 2000, 2005 and 2010. Data on migration from the World Bank. Migration normalized by product of bordering country populations and to have mean zero and unit variance. Observation unit is a country-border piece-year. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

Table 2.A4

Permeability: Tariffs							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.	Log Lights	Log Lights	Growth, Lights	Growth, Lights	Growth, Lights	Growth, Lights	Growth, Lights
Nat. Output / Growth	.63***	.60*	1.47***	1.05**	.60*	.90**	.59*
	(.24)	(.31)	(.30)	(.45)	(.35)	(.36)	(.34)
Nat. Output / Growth X Barriers	.11	.06	-.66**	-.34	-.33	-.36	-.34
	(.22)	(.26)	(.26)	(.34)	(.30)	(.31)	(.30)
Rule of Law					2.76***		2.44***
					(.98)		(.86)
Rule of Law X Barriers					.00		-.07
					(.82)		(.75)
Fraction of Roads Paved						1.99*	.75
						(1.17)	(.97)
Fraction of Roads Paved X Barriers						.43	.49
						(.91)	(.85)
Output Source	Lights	GDP	Lights	GDP	GDP	GDP	GDP
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Observations	6620	6620	1324	1324	1316	1310	1310
No. Borders	261	261	261	261	258	256	256
R^2	.09	.04	.21	.05	.11	.08	.11

(2.A4)

Data on lights and population available from NOAA and CIESIN, respectively, for 1990, 1995, 2000, 2005 and 2010. Data on tariff barriers from the World Bank. Tariff barriers normalized to have mean zero and unit variance. Observation unit is a country-border piece-year. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

Table 2.A5

Trade Channel: Genetic Distance							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.	Log Lights	Log Lights	Growth, Lights	Growth, Lights	Growth, Lights	Growth, Lights	Growth, Lights
Nat. Output / Growth	.59***	.53**	1.29***	.90***	.47*	.77***	.45*
	(.20)	(.24)	(.28)	(.34)	(.25)	(.26)	(.25)
Nat. Output / Growth X Gen. Dist.	-.24**	-.31*	-.25	.22	.33**	.24	.32**
	(.11)	(.17)	(.21)	(.18)	(.15)	(.16)	(.16)
Rule of Law					2.69***		2.52***
					(.94)		(.77)
Rule of Law X Gen. Dist.					-.32		-.31
					(.55)		(.58)
Fraction of Roads Paved						1.80	.58
						(1.17)	(1.07)
Fraction of Roads Paved X Gen. Dist.						-.55	.19
						(.52)	(.51)
Output Source	Lights	GDP	Lights	GDP	GDP	GDP	GDP
Border-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. Observations	6760	6760	1352	1352	1342	1336	1336
No. Borders	270	270	270	270	266	264	264
R^2	.11	.06	.18	.05	.11	.08	.11

(2.A5)

Data on lights and population available from NOAA and CIESIN, respectively, for 1990, 1995, 2000, 2005 and 2010. Data on genetic distance is from Spolaore and Wacziarg. Genetic distance normalized to have mean zero and unit variance. Observation unit is a country-border piece-year. Robust standard errors clustered on border and taking into account infill asymptotics in parentheses. Each observation weighted by population in the respective buffer piece.

2.10 Appendix II: Proof of Proposition

2.10.1 Setup

I consider the properties of the standard local polynomial estimator in an infill asymptotics context. The core result is that the local polynomial estimator has smaller standard errors if the random shocks of the stochastic process are correlated than if they are not, and that a feasible and consistent estimator exists to estimate these standard errors.

To present the results formally, define the following notation: let

$$G = \left\{ \frac{u}{N} \right\}_{u=-N}^{N}$$

be a sequence with resolution N. Let s denote a generic element of G. Note that the domain of G remains bounded, and is contained in $[-1, 1]$, which is an arbitrary interval in R up to a normalization. The process is assumed to have a discontinuity at $s = 0$. In particular, we are interested in using the sequence

$$\left\{ y\left(\frac{u}{N}\right) \right\}_{u=-N}^{N}$$

where $y(s)$ is a scalar, to predict the value

$$y_{+}(0) := \lim_{u \downarrow 0} y(u)$$

The value y_{-} is defined analogously:

$$y_{-}(0) := \lim_{u \uparrow 0} y(u)$$

The discontinuity at 0 is defined to be

$$\bar{\Delta} = y_{+}(0) - y_{-}(0)$$

Estimating $y_{+}(0)$ and $y_{-}(0)$ is of central interest in this document.

Define the local polynomial estimator of $y_{+}(0)$ by

$$\hat{\alpha}_N^{+} = e_1' \left(\sum_{u=1}^{N} \left(\frac{1}{Nh}\right) k\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right)' \right)^{-1} \left(\sum_{u=1}^{N} \left(\frac{1}{Nh}\right) k\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right) y\left(\frac{u}{N}\right) \right)$$

where $X()$ is a $K \times 1$ vector of polynomials, and $k\left(\frac{u}{Nh}\right)$ is a positive kernel. Define the local polynomial estimator of $y_{-}(0)$ similarly by $\hat{\alpha}_N^{-}$. For convenience, define $D_N = \sum_{u=1}^{N} \left(\frac{1}{Nh}\right) k\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right)'$, the denominator of this expression and note that it is nonstochastic and converges to $D = \int_0^{\infty} k(u) X(u) X(u)' du$. Then, the local polynomial estimator of the discontinuity at zero is given by

$$\hat{\Delta}_N = \hat{\alpha}_N^{+} - \hat{\alpha}_N^{-}$$

Finally, I define the *increment* of a stochastic process $y(s)$ to be given by

$$y_{N,s} = y(s) - y\left(s - \frac{1}{N}\right)$$

Assumptions on the Bandwidth and Kernel

B1. Define the bandwidth $h(N)$. Then, $\lim_{N\to\infty} h(N) = 0$, $\lim_{N\to\infty} Nh(N) = \infty$ and $\lim_{N\to\infty} Nh(N)^2 = 0$

B2. The kernel $k()$ satisfies $\int_0^{\infty} \left(\int_u^{\infty} k(v) v^p dv \right) du < \infty$ for all $p \leq K$.

Assumptions on $y(s)$

We assume that we can decompose $y(s)$ as

$$y(s) = F(s) + v(s) + e(s)$$

Consider some assumptions on the components of $y(s)$:

1. $F(s) \in C^1[0,1]$ is a deterministic function.

2. $v(s)$ is a random shock that is independent across realizations: $E(v(s)) = 0$, $E(v(s)v(t)) = \bar{V}(s) \cdot 1(s=t)$, with $\bar{V}(s) \in C[0,1]$, and $E\left(v(s)^{2+\delta}\right) \leq K < \infty$.

2' $v(s)$ is identically zero.

3. $e(s)$ is a random shock such that $E(e(s)) = 0$, $E(e(s)v(t)) = 0$, but

$$E(e(s)e(t)) = C(s,t)$$

for some function $C(s,t)$ that a) belongs to $C^2\left\{(s,t) \in [0,1]^2 : s \neq t\right\}$, b) satisfies $\lim_{t\to s^+} C(s,t) = \lim_{t\to s^-} C(s,t)$, and c) satisfies $V(s) = C(s,s) \in C^1([0,1])$. Moreover,

$$\operatorname{cov}\left(Ne_{N,u}^2, Ne_{N,u'}^2\right) = O\left(\operatorname{cov}\left(\sqrt{N}e_{N,u}, \sqrt{N}e_{N,u'}\right)^2\right)$$

3' $e(s)$ is defined as in Assumption 3, but with an additional assumption.

$$\sigma(s) := V'(s) - 2\lim_{N\to\infty} C_1\left(s, s-\frac{1}{N}\right) \geq 0,\ \forall s \in [0,1]$$

4. The increments of the error process $e_{N,u} = e\left(\frac{u}{N}+\frac{1}{N}\right) - e\left(\frac{u}{N}\right)$ are associated if N is large enough.

It is insightful to compare these assumptions with standard assumptions from the geostatistics literature, which deals with spatially correlated processes. A very general assumption in that literature is that the zero-mean stochastic process $y(s)$ has a stationary variogram, which is continuous everywhere except at zero. Hence,

$$E(y(s)y(t)) = \frac{1}{2}[V(s) + V(t)] - \gamma(|s-t|)$$

where γ is a continuous function with $\gamma(0) = 0$ and $\lim_{s\to 0}\gamma(s) =: c_0 > 0$. Defining $\tilde{\gamma}(s) = \gamma(s) - c_0$, we can then write

$$E(y(s)y(t)) = \frac{1}{2}[V(s) + V(t)] - \tilde{\gamma}(|s-t|) + c_0 \cdot (s=t)$$

Now, under Assumptions 1, 2 and 3, we have

$$E((v(s)+e(s))(v(t)+e(t))) = C(s,t) + \bar{V}(s)\cdot(s=t)$$

where $C(s,t)$ is any positive definite, continuous function satisfying some smoothness conditions, and $\bar{V}(s)$ is also a continuous function. Hence, any sufficiently smooth (except at zero) variogram and variance function satisfy these assumptions. It is immediate that any random field with $c_0 = 0$ satisfies assumptions 1, 2' and 3. Moreover, the geostatistics literature tends to work with Gaussian stochastic processes (or Gaussian random fields), and it is known that jointly Gaussian random variables x and y satisfy $\operatorname{cov}(x^2, y^2) = 2\operatorname{cov}(x,y)^2$ by Isserlis's Theorem. Therefore, any Gaussian random field with a sufficiently smooth variogram and variance function satisfies Assumptions 1, 2 and 3'. Therefore, the set of assumptions considered is extremely general.

2.10.2 Propositions

We then have the following propositions (with $V(0)$ being the variance of $e(0)$ and $\bar{V}(0)$ being the variance of $v(0)$)

Proposition 1 *Suppose that Assumptions 1, 2 and 3 hold. Then, $\hat{\alpha}_N^+$ consistently estimates $y_+(0) - v_+(0)$, and*

$$\begin{aligned}\sqrt{Nh}\left(\hat{\alpha}_N^+ - (y_+(0) - v_+(0))\right) &\to {}^d N(0, V_1) \\ V_1 \quad : &= \bar{V}(0)e_1'D^{-1}\left(\int_0^\infty (k(u))^2 X(u)X(u)'du\right)D^{-1}e_1\end{aligned}$$

Moreover, a feasible and consistent estimator of V_1 is

$$\hat{V}_{1,N} = \frac{1}{2} e_1' D_N^{-1} \left[\sum_{u=1}^{N} \hat{e}_{N,u}^2 \frac{1}{Nh} k^2 \left(\frac{u}{Nh} \right) X \left(\frac{u}{Nh} \right) X \left(\frac{u}{Nh} \right)' \right] D_N^{-1} e_1$$

Proposition 2 *Suppose that Assumptions 1, 2 and 3 hold. Then the White estimator of the variance of $\hat{\alpha}_N^+$ asymptotically overestimates the true variance in expectation. Specifically,*

$$NhE \left(\hat{V}_{OLS} \right) \to \left(\bar{V}(0) + V(0) \right) e_1' D^{-1} \left(\int_0^\infty (k(u))^2 X(u) X(u)' du \right) D^{-1} e_1$$

Proposition 3 *Suppose that Assumptions 1, 2', 3' and 4 hold. Then, $\hat{\alpha}_N^+$ consistently estimates $y_+(0)$ and*

$$\begin{aligned}
\left(\sqrt{h} \right)^{-1} \left(\alpha_N^+ - y_+(0) \right) &\to {}^d N(0, V_2) \\
V_2 \quad : &= \sigma(0) e_1' D^{-1} \left[\left(\int_0^\infty \left(\int_u^\infty k(v) X(v) dv \right) \left(\int_u^\infty k(v) X(v) dv \right)' du \right) \right] D^{-1} e_1
\end{aligned}$$

Moreover, a feasible and consistent estimator of V_2 is

$$\begin{aligned}
\hat{V}_2 &= \frac{1}{2} e_1' D_N^{-1} \left[\sum_{u=1}^{N} \frac{1}{Nh} k \left(\frac{u}{Nh} \right) X \left(\frac{u}{Nh} \right) y_{N,u}^2 \right] \times \\
& e_1' D_N^{-1} \left[\frac{1}{h(N)} \sum_{u=1}^{N} \left(\sum_{v=u}^{N} \frac{1}{Nh} k \left(\frac{v}{Nh} \right) X \left(\frac{v}{Nh} \right) \right) \left(\sum_{v=u}^{N} \frac{1}{Nh} k \left(\frac{v}{Nh} \right) X \left(\frac{v}{Nh} \right) \right)' \right] D_N^{-1} e_1
\end{aligned}$$

2.10.3 Proof of Proposition 1

Consistency

Suppose that Assumptions 1, 2 and 3 hold. Consider

$$\begin{aligned}
Z_N^+ &= \alpha_N^+ - y_+(0) \\
&= e_1' D_N^{-1} \left(\sum_{u=1}^{N} \left(\frac{1}{Nh} \right) k \left(\frac{u}{Nh} \right) X \left(\frac{u}{Nh} \right) \left(F \left(\frac{u}{N} \right) - F_+(0) \right) \right) \\
&\quad + e_1' D_N^{-1} \left(\sum_{u=1}^{N} \left(\frac{1}{Nh} \right) k \left(\frac{u}{Nh} \right) X \left(\frac{u}{Nh} \right) \left(v \left(\frac{u}{N} \right) - v_+(0) \right) \right) \\
&\quad + e_1' D_N^{-1} \left(\sum_{u=1}^{N} \left(\frac{1}{Nh} \right) k \left(\frac{u}{Nh} \right) X \left(\frac{u}{Nh} \right) \left(e \left(\frac{u}{N} \right) - e_+(0) \right) \right)
\end{aligned}$$

I first show that α_N^+ is asymptotically unbiased: $\lim_{N \to \infty} E\left(Z_N^+\right) = 0$. To do this, I present a very simple lemma that will be useful in further analysis.

Lemma 4 : *Suppose $\lim_{N \to \infty} \sum_{u=1}^{N} \left(\frac{1}{Nh} \right) S \left(\frac{u}{Nh} \right) = \int_0^\infty S(u) du$ exists in R, and $F(x)$ is a continuous function on $[0,1]$. Then,*

$$\lim_{N \to \infty} \sum_{u=1}^{N} \left(\frac{1}{Nh} \right) S \left(\frac{u}{Nh} \right) F \left(\frac{u}{N} \right) = F(0) \int_0^\infty S(u) du$$

Proof.

$$\begin{aligned}
\sum_{u=1}^{N} \left(\frac{1}{Nh} \right) S \left(\frac{u}{Nh} \right) \left| F \left(\frac{u}{N} \right) - F(0) \right| &\leq \sup_{u \in [0,\tau]} |F(u) - F(0)| \sum_{u=1}^{N} \left(\frac{1}{Nh} \right) S \left(\frac{u}{Nh} \right) \\
&\quad + \sup_{u \in [0,1]} |F(u) - F(0)| \sum_{u=\tau N}^{\tau N} \left(\frac{1}{Nh} \right) S \left(\frac{u}{Nh} \right)
\end{aligned}$$

Since F is continuous, we have $\lim_{\tau\to 0}\sup_{u\in[0,\tau]}|F(u)-F(0)|=0$. Moreover, since $\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)S\left(\frac{u}{Nh}\right)$ is a Riemann sum, it is immediate that

$$\lim_{N\to\infty}\sum_{u=1}^{\tau N}\left(\frac{1}{Nh}\right)S\left(\frac{u}{Nh}\right)=\int_0^\infty S(u)\,du \text{ and } \lim_{N\to\infty}\sum_{u=\tau N}^{N}\left(\frac{1}{Nh}\right)S\left(\frac{u}{Nh}\right)=0,\ \forall\tau>0$$

Therefore, for any ε, we can find a $\tau(\varepsilon)$ small enough that $\sup_{u\in[0,\tau(\varepsilon)]}|F(u)-F(0)|<\varepsilon$, and an N large enough that $\sum_{u=1}^{\tau(\varepsilon)N}\left(\frac{1}{Nh}\right)S\left(\frac{u}{Nh}\right)=(1-\varepsilon)\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)S\left(\frac{u}{Nh}\right)$.

Hence, $\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)S\left(\frac{u}{Nh}\right)\left|F\left(\frac{u}{N}\right)-F(0)\right|\le\varepsilon\left[(1-\varepsilon)+\sup_{u\in[0,1]}|F(u)-F(0)|\right]$, and taking the limit as ε goes to zero completes the proof. ■

To prove $\lim_{N\to\infty}E\left(Z_N^+\right)=0$, I apply Lemma 4 (because F is continuous):

$$\lim_{N\to\infty}E\left(Z_N^+\right)=\lim_{N\to\infty}e_1'D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\left(F\left(\frac{u}{N}\right)-F_+(0)\right)\right)=0$$

Now, I show that $\lim_{N\to\infty}\operatorname{var}\left(Z_N^++v_+(0)\right)=0$

$$\begin{aligned}\operatorname{var}\left(Z_N^++v_+(0)\right) &= E\left(\left(Z_N^++v_+(0)\right)^2\right)-E\left(Z_N^++v_+(0)\right)^2\\ &= E\left(\left[e_1'D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)v\left(\frac{u}{N}\right)\right)\right]^2\right)\quad\text{(I)}\\ &\quad +E\left(\left[e_1'D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\left(e\left(\frac{u}{N}\right)-e_+(0)\right)\right)\right]^2\right)\quad\text{(II)}\end{aligned}$$

Consider term (I). We have

$$E\left(\left[e_1'D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)v\left(\frac{u}{N}\right)\right)\right]^2\right)=\frac{1}{Nh}e_1'D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k^2\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)'\bar{V}\left(\frac{u}{N}\right)\right)D_N^{-1}e_1$$

where $\bar{V}\left(\frac{u}{N}\right)=E\left(v\left(\frac{u}{N}\right)^2\right)$. Now, since $\lim_{N\to\infty}\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k^2\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)'=\int_0^\infty k^2(u)X(u)X(u)'du$ and $\bar{V}(u)$ is continuous, we can appeal to Lemma 4 to argue that

$$\lim_{N\to\infty}NhE\left(\left[e_1'D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)v\left(\frac{u}{N}\right)\right)\right]^2\right)\to\bar{V}_+(0)\,e_1'D^{-1}\left(\int_0^\infty k^2(u)X(u)X(u)'du\right)D^{-1}e_1=:V_1$$

because

$$\lim_{N\to\infty}\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k^2\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)'\left|\bar{V}\left(\frac{u}{N}\right)-\bar{V}_+(0)\right|=0$$

Now, consider term (II). We can rewrite it in terms of increments $e_{N,u}=e\left(\frac{u}{N}\right)-e\left(\frac{u}{N}-\frac{1}{N}\right)$ as the following:

$$\text{(II)}=\left(e_1'D_N^{-1}\sum_{u=1}^{N}\left(\sum_{v=u}^{N}\frac{1}{Nh}k\left(\frac{v}{Nh}\right)X\left(\frac{v}{Nh}\right)\right)e_{N,u}\right)^2$$

by noting that

$$e\left(\frac{v}{N}\right)-e_+(0)=\sum_{u=1}^{v}\left(e\left(\frac{u}{N}\right)-e\left(\frac{u}{N}-\frac{1}{N}\right)\right)=\sum_{u=1}^{v}e_{N,u}$$

We can further break down term (II) by noting that

$$\begin{aligned}
\text{(II)} &= e_1' D_N^{-1} \sum_{u=1}^{N} \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right)' E\left(e_{N,u}^2\right) D_N^{-1} e_1 \quad (1) \\
&+ e_1' D_N^{-1} \sum_{u=1}^{N} \sum_{\substack{u'=1 \\ u' \neq u}}^{N} \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \left(\sum_{v=u'}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right)' E\left(e_{N,u} e_{N,u'}\right) D_N^{-1} e_1 \quad (2)
\end{aligned}$$

Now, by Assumption 3, we have

$$\begin{aligned}
E\left(e_{N,u} e_{N,u'}\right) &= C(s,t) - C\left(s, t - \frac{1}{N}\right) - C\left(s - \frac{1}{N}, t\right) + C\left(s - \frac{1}{N}, t - \frac{1}{N}\right) \\
&= \left(C(s,t) - C\left(s, t - \frac{1}{N}\right)\right) - \left(C\left(s - \frac{1}{N}, t\right) - C\left(s - \frac{1}{N}, t - \frac{1}{N}\right)\right) \\
&\rightarrow \frac{1}{N^2} \frac{\partial^2}{\partial s \partial t} C(s,t) = O\left(\frac{1}{N^2}\right)
\end{aligned}$$

and

$$\begin{aligned}
E\left(e_{N,u}^2\right) &= V(s) - 2C\left(s, s - \frac{1}{N}\right) + V\left(s - \frac{1}{N}\right) \\
&= \left(V(s) - V\left(s - \frac{1}{N}\right)\right) - 2\left(C\left(s, s - \frac{1}{N}\right) - C\left(s - \frac{1}{N}, s - \frac{1}{N}\right)\right) \\
&= \frac{1}{N}\left[\Delta V(s) - 2\Delta_1 C\left(s, s - \frac{1}{N}\right)\right] \rightarrow \frac{1}{N}\left[V'(s) - 2 \lim_{N \to \infty} C_1\left(s, s - \frac{1}{N}\right)\right] = O\left(\frac{1}{N}\right)
\end{aligned}$$

if $C(s,t)$ is not twice differentiable at $s = t$ with uniformly bounded second derivatives. (If it is twice differentiable at $s = t$, then $E\left(e_{N,u}^2\right) = O\left(\frac{1}{N^2}\right)$).

Under the boundedness conditions on $\frac{\partial^2}{\partial s \partial t} C(s,t)$, $C_1\left(s, s - \frac{1}{N}\right)$ and $V'(s)$, the convergence can be taken as uniform, and

$$\sup_{(s,t) \in [0,1]^2} \frac{\partial^2}{\partial s \partial t} C(s,t) \leq K_2 < \infty$$

and

$$\sup_{(s,t) \in [0,1]^2} \left[V'(s) - 2 \lim_{N \to \infty} C_1\left(s, s - \frac{1}{N}\right)\right] \leq K_1 < \infty$$

Therefore,

$$\begin{aligned}
\limsup_{N \to \infty} \text{(II)} &\leq hK_1 e_1' D^{-1} \left(\int_0^\infty \left(\int_u^\infty k(v) X(v) \right) \left(\int_u^\infty k(v) X(v) \right)' du \right) D^{-1} e_1 \\
&+ h^2 K_2 e_1' D^{-1} \left[\left(\int_0^\infty \left(\int_u^\infty k(v) X(v) \right) du \right) \left(\int_0^\infty \left(\int_u^\infty k(v) X(v) \right) du \right)' \right] D^{-1} e_1
\end{aligned}$$

and hence,

$$\text{(II)} = O(h)$$

Therefore,

$$\text{var}\left(Z_N^+ + v_+(0)\right) = \frac{1}{Nh} V_1 + O\left(Nh^2\right) \rightarrow 0 \text{ as } N \rightarrow \infty$$

and consistency is proved.

Asymptotic Distribution

To find the asymptotic distribution of α_N^+, we note that

$$\sqrt{Nh}\left(\alpha_N^+ - \left(y_+\left(0\right) - v_+\left(0\right)\right)\right) = \sqrt{Nh}e_1' D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\left(F\left(\frac{u}{N}\right) - F_+\left(0\right)\right)\right) \quad (1)$$
$$+\sqrt{Nh}e_1' D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)v\left(\frac{u}{N}\right)\right) \quad (2)$$
$$+\sqrt{Nh}e_1' D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\left(e\left(\frac{u}{N}\right) - e_+\left(0\right)\right)\right) \quad (3)$$

By decomposing Term 1 into increments as in the proof of consistency, we see that

$$(1) = \sqrt{Nh}e_1' D_N^{-1}\sum_{u=1}^{N}\left(\sum_{v=u}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)F_{N,u}$$

Since $F\left(s\right)$ is continuously differentiable by Assumption 1, we have that $F_{N,u} = \frac{1}{N}\Delta F_{N,u} \to \frac{1}{N}F'\left(\frac{u}{N}\right)$. Under the boundedness conditions of Assumption 1, we therefore have

$$\sup_{s\in[0,1]} F'\left(s\right) \leq K < \infty$$

Therefore,

$$\begin{aligned}(1) &\leq K\sqrt{Nh}e_1' D_N^{-1}\sum_{u=1}^{N}\frac{1}{N}\left(\sum_{v=u}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\\ &= Kh\sqrt{Nh}e_1' D_N^{-1}\sum_{u=1}^{N}\frac{1}{Nh}\left(\sum_{v=u}^{N}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\end{aligned}$$

and

$$(1) \to K\left(\lim_{N\to\infty} h\sqrt{Nh}\right)e_1' D^{-1}\left(\int_0^{\infty}\left(\int_u^{\infty}k\left(v\right)X\left(v\right)dv\right)du\right)$$

Hence, if $\lim_{N\to\infty} h\sqrt{Nh} = 0$, the asymptotic bias is zero. This is implied by the bandwidth assumption $\lim_{N\to\infty} Nh^2 = 0$.

By the variance results from the consistency proof, term (2) is $O_p\left(1\right)$, while term (3) is $o_p\left(1\right)$. Therefore, we are interested only in the asymptotic distribution of term (2). Since $v\left(s\right)$ is independent of $v\left(t\right)$ for all t and s, term (2) is a sum of independent random variables. To satisfy the hypotheses of the Liapunov Central Limit Theorem, we must prove that

$$\lim_{N\to\infty}\sum_{u=1}^{N}\operatorname{var}\left(\sqrt{Nh}e_1' D_N^{-1}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)v\left(\frac{u}{N}\right)\right) < \infty$$

and

$$\lim_{N\to\infty}\sum_{u=1}^{N}E\left[\left(\sqrt{Nh}e_1' D_N^{-1}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)v\left(\frac{u}{N}\right)\right)^{2+\delta}\right] = 0 \; \exists\delta > 0$$

The first condition follows from the computation of the variance. The second condition follows because

$$\begin{aligned}
&\lim_{N\to\infty}\sum_{u=1}^{N}E\left[\left(\sqrt{Nh}e_1'D_N^{-1}\left(\frac{1}{Nh}\right)k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)v\left(\frac{u}{N}\right)\right)^{2+\delta}\right]\\
=\ &\lim_{N\to\infty}\frac{1}{(Nh)^{\delta/2}}\sum_{u=1}^{N}\left(\frac{1}{Nh}\right)\left[\left(e_1'D_N^{-1}X\left(\frac{u}{Nh}\right)\right)^{2+\delta}\right]k^{2+\delta}\left(\frac{u}{Nh}\right)E\left(v^{2+\delta}\left(\frac{u}{N}\right)\right)\\
=\ &\lim_{N\to\infty}\frac{1}{(Nh)^{\delta/2}}\cdot\int_0^{\infty}\left[\left(e_1'D^{-1}X(u)\right)^{2+\delta}\right]k^{2+\delta}(u)\,E\left(v^{2+\delta}(u)\right)du=0
\end{aligned}$$

since $E\left(v^{2+\delta}(u)\right)\leq K<\infty$. Therefore, the hypothesis of the Liapunov CLT are satisfied, and we have

$$\sqrt{Nh}\left(\hat{\alpha}_N^{+}-\left(y_{+}(0)-v_{+}(0)\right)\right)\to^{d}N(0,V_1)$$

as desired.

Feasible Estimation

Finally, I show that a feasible and consistent estimator of V_1 is

$$\hat{V}_{1,N}=\frac{1}{2}e_1'D_N^{-1}\left[\sum_{u=1}^{N}\hat{e}_{N,u}^2\frac{1}{Nh}k^2\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)'\right]D_N^{-1}e_1$$

which replaces the residuals in the OLS estimator with the residual increments $e_{N,u}^2$.

To prove this, we show that $\hat{V}_{1,N}^{A}=Nh\frac{1}{2}\sum_{u=1}^{N}\hat{e}_{N,u}^2\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'$ is a consistent estimator of $V_{1,N}^{A}=\bar{V}(0)\int_0^{\infty}k^2(u)X(u)X(u)'du$

First, we see that $\hat{V}_{1,N}^{A}$ is asymptotically unbiased for $V_{1,N}^{A}$:

$$\begin{aligned}
E\left(\hat{V}_{1,N}^{A}\right)&=\frac{1}{2}\sum_{u=1}^{N}\frac{1}{Nh}\left(k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'E\left(\hat{e}_{N,u}^2\right)\\
&=\frac{1}{2}e_1'D_N^{-1}\left[\left(\begin{array}{c}
\sum_{u=1}^{N}\frac{1}{Nh}\left(k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'\times\\
E\left(v\left(\frac{u}{N}\right)-v\left(\frac{u-1}{N}\right)\right)^2+E\left(\hat{e}_{N,u}^2\right)\\
+E\left(\left(\frac{1}{N}\Delta F_{N,u}-\frac{1}{N}\Delta X\left(\frac{u}{N}\right)'E\left(\hat{\beta}\right)+\frac{1}{N}\Delta X\left(\frac{u}{N}\right)'\left(E\left(\hat{\beta}\right)-\beta\right)\right)^2\right)\\
+E\left(\left(\frac{1}{N}\Delta F_{N,u}-\frac{1}{N}\Delta X\left(\frac{u}{N}\right)'E\left(\hat{\beta}\right)+\frac{1}{N}\Delta X\left(\frac{u}{N}\right)'\left(E\left(\hat{\beta}\right)-\beta\right)\right)\left(v_{N,u}+e_{N,u}\right)\right)
\end{array}\right)\right]\\
&=\frac{1}{2}e_1'D_N^{-1}\left[\sum_{u=1}^{N}\frac{1}{Nh}\left(k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'\left(\left(\bar{V}\left(\frac{u}{N}\right)+\bar{V}\left(\frac{u-1}{N}\right)\right)+o(1)\right)\right]
\end{aligned}$$

Now, invoking Lemma 4, it is clear that $\lim_{N\to\infty}\left[\sum_{u=1}^{N}\frac{1}{Nh}\left(k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'\bar{V}\left(\frac{u}{N}\right)\right]=\bar{V}(0)\int_0^{\infty}k^2(u)X(u)X(u)'$ since $\bar{V}\left(\frac{u}{N}\right)$ is continuous. The analysis for $\sum_{u=1}^{N}\frac{1}{Nh}\left(k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'\bar{V}\left(\frac{u}{N}\right)$ is similar. Hence, $E\left(\hat{V}_{1,N}\right)\to V_1$.

To compute the variance of $\hat{V}_{1,N}^{A}$, note that

$$\begin{aligned}
E\left(\hat{e}_{N,u}^2\hat{e}_{N,u'}^2\right)&=E\left(v_{N,u}^2v_{N,u'}^2\right)+E\left(e_{N,u}^2e_{N,u}^2\right)+E\left(v_{N,u}^2\right)E\left(e_{N,u'}^2\right)+E\left(v_{N,u'}^2\right)E\left(e_{N,u}^2\right)+O\left(\frac{1}{N^2}\right)\\
&=E\left(v_{N,u}^2v_{N,u'}^2\right)+o(1)
\end{aligned}$$

Moreover,

$$E\left(v_{N,u}^2 v_{N,u'}^2\right) - E\left(v_{N,u}^2\right) E\left(v_{N,u'}^2\right) = \begin{cases} \left[E\left(v\left(\frac{u-1}{N}\right)^4\right) - \left(E\left(v\left(\frac{u-1}{N}\right)^2\right)\right)^2\right] =: S_1\left(\frac{u}{N}\right) \text{ if } |u-u'| = 1 \\ \left[\begin{array}{c} \left(E\left(v\left(\frac{u}{N}\right)^4\right) - \left(E\left(v\left(\frac{u}{N}\right)^2\right)\right)^2\right) \\ +\left(E\left(v\left(\frac{u-1}{N}\right)^4\right) - \left(E\left(v\left(\frac{u-1}{N}\right)^2\right)\right)^2\right) \\ +4E\left(v\left(\frac{u}{N}\right)^2\right) E\left(v\left(\frac{u-1}{N}\right)^2\right) \end{array}\right] =: S_0\left(\frac{u}{N}\right) \text{ if } u = u' \\ 0, \text{ if } |u-u'| > 1 \end{cases}$$

Therefore,

$$\begin{aligned} (Nh)^2 E\left(\left(\hat{V}_{1,N}^A\right)^2\right) &= \left[\sum_{u=1}^N \sum_{u'=1}^N \left(\frac{1}{Nh}\right)^2 k^2\left(\frac{u}{Nh}\right) k^2\left(\frac{u'}{Nh}\right) X\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right)' X\left(\frac{u'}{Nh}\right) X\left(\frac{u'}{Nh}\right)' E\left(e_{N,u}^2 e_{N,u'}^2\right)\right] \\ &= \left[\sum_{u=1}^N \sum_{u'=1}^N \left(\frac{1}{Nh}\right)^2 k^2\left(\frac{u}{Nh}\right) k^2\left(\frac{u'}{Nh}\right) X\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right)' X\left(\frac{u'}{Nh}\right) X\left(\frac{u'}{Nh}\right)' E\left(e_{N,u}^2 e_{N,u'}^2\right)\right] + O\left(\frac{1}{N}\right) \\ &= \left[E\left(\hat{V}_{1,N}^A\right)\right]^2 + O\left(\frac{1}{N}\right) + \left[\sum_{u=1}^N \left(\frac{1}{Nh}\right)^2 k^4\left(\frac{u}{Nh}\right) \left(X\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right)'\right)^2 S_0\left(\frac{u}{N}\right)\right] \\ &\quad + \left[\sum_{u=1}^N \left(\frac{1}{Nh}\right)^2 k^4\left(\frac{u}{Nh}\right) \left(X\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right)'\right)^2 S_1\left(\frac{u}{N}\right)\right] \end{aligned}$$

Finally, note that

$$\begin{aligned} &\sum_{u=1}^N \left(\frac{1}{Nh}\right)^2 k^4\left(\frac{u}{Nh}\right) \left(X\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right)'\right)^2 S_0\left(\frac{u}{N}\right) \\ \leq &\left\{\sup_u S\left(\frac{u}{N}\right)\right\} \frac{1}{Nh} \sum_{u=1}^N \left(\frac{1}{Nh}\right) k^4\left(\frac{u}{Nh}\right) \left(X\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right)'\right)^2 \to 0. \end{aligned}$$

and similarly for $\sum_{u=1}^N \left(\frac{1}{Nh}\right)^2 k^4\left(\frac{u}{Nh}\right) \left(X\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right)'\right)^2 S_1\left(\frac{u}{N}\right)$. Therefore,

$$E\left(\hat{V}_{1,N}^A\right) \to \left[E\left(V_1^A\right)\right]^2$$

A straightforward application of the Slutsky theorem is sufficient to show that $\hat{V}_{1,N}$ is a consistent estimator of V_1.

2.10.4 Proof of Proposition 2

The White estimator of the variance of the local polynomial estimator is given by

$$V_N^{OLS} = e_1 D_N^{-1} \left(\sum_{u=1}^N \hat{e}^2\left(\frac{u}{N}\right) \frac{1}{Nh} k^2\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right)'\right) D_N^{-1} e_1$$

where

$$\hat{e}^2(u) = \left(y\left(\frac{u}{N}\right) - X\left(\frac{u}{N}\right)' \hat{\beta}\right)^2$$

and $\hat{\beta}$ is the local polynomial estimator of the entire vector of derivatives of $y(s)$ at zero. Therefore, $X(0)'\hat{\beta} = \alpha_N^+$. Note from the previous proof that $\hat{\beta}$ is a consistent estimator of its expected value, and in particular, that $\sqrt{Nh}\left(\hat{\beta} - E\left(\hat{\beta}\right)\right) = O_p(1)$.

Therefore,

$$\hat{e}(u)=\left(\left(F\left(\frac{u}{N}\right)-X\left(\frac{u}{N}\right)'E\left(\hat{\beta}\right)+v\left(\frac{u}{N}\right)+e\left(\frac{u}{N}\right)\right)+X\left(\frac{u}{N}\right)'\left(\hat{\beta}-E\left(\hat{\beta}\right)\right)\right)$$

and

$$\begin{aligned}
E\left(V_N^{OLS}\right) &= e_1 D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'\bar{V}\left(\frac{u}{N}\right)\right)D_N^{-1}e_1 \quad (1)\\
&+e_1 D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'V\left(\frac{u}{N}\right)\right)D_N^{-1}e_1 \quad (2)\\
&+e_1 D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'\left(F\left(\frac{u}{N}\right)-X\left(\frac{u}{N}\right)'E\left(\hat{\beta}_N\right)\right)^2\right)D_N^{-1}e_1 \quad (3)\\
&+2e_1 D_N^{-1}\left(\begin{array}{c}\sum_{u=1}^{N}\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'X\left(\frac{u}{N}\right)'\left(\hat{\beta}_N-E\left(\hat{\beta}_N\right)\right)\times\\ \left(F\left(\frac{u}{N}\right)-X\left(\frac{u}{N}\right)'E\left(\hat{\beta}_N\right)+v\left(\frac{u}{N}\right)+e\left(\frac{u}{N}\right)\right)\end{array}\right)D_N^{-1}e_1 \quad (4)\\
&+e_1 D_N^{-1}\left(\sum_{u=1}^{N}\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)'X\left(\frac{u}{N}\right)'\operatorname{var}\left(\hat{\beta}_N\right)X\left(\frac{u}{N}\right)\right)D_N^{-1}e_1 \quad (5)
\end{aligned}$$

It is clear that $Nh\cdot(1)$ and $Nh\cdot(2)$ converge to

$$\bar{V}(0)\,e_1 D^{-1}\left(\int_0^{\infty}k^2(u)\,X(u)\,X(u)'\,du\right)D^{-1}e_1$$

and

$$V(0)\,e_1 D^{-1}\left(\int_0^{\infty}k^2(u)\,X(u)\,X(u)'\,du\right)D^{-1}e_1$$

respectively. Since $\lim_{N\to\infty}X\left(\frac{u}{N}\right)'E\left(\hat{\beta}_N\right)=F_+(0)$, we have that $Nh\cdot(3)\to 0$. Now, we know from the proof of Proposition 1 that $\sqrt{Nh}\left(\hat{\beta}_N-E\left(\hat{\beta}_N\right)\right)$ converges to an $O_p(1)$ random variable, and therefore, so does $(Nh)^{3/2}\cdot(4)$. Hence, $Nh\cdot(4)\to 0$. Finally, we know that $Nh\operatorname{var}\left(\hat{\beta}_N\right)\to O(1)$, so $(Nh)^2\cdot(5)\to O(1)$ and hence, $Nh\cdot(5)\to 0$.

Therefore,

$$NhV^{OLS}\to\left(\bar{V}(0)+V(0)\right)e_1 D^{-1}\left(\int_0^{\infty}k^2(u)\,X(u)\,X(u)'\,du\right)D^{-1}e_1$$

which is strictly larger than the true asymptotic variance of the local polynomial estimator. Intuitively, the White estimator assumes that the errors around the trend are independent, and fails to account for the fact that the local polynomial estimator exploits correlations between errors to improve its predictive power. Only the "independent" part of the error contributes to the asymptotic variance; the "correlated" part of the error can ultimately be perfectly predicted as the resolution of the data becomes infinite.

2.10.5 Proof of Proposition 3

For this proof, we assume that $v(s)=0$ for all s

Consistency

We have already shown that α_N^+ is consistent for $y_+(0)-v_+(0)$ in proposition 1. It therefore remains to compute its asymptotic variance.

Under assumption 2', the leading term of the variance in Proposition 1 is zero. Therefore, we consider term (II) from the proof of Proposition 1:

$$
\begin{aligned}
(\mathrm{II}) &= e_1' D_N^{-1} \sum_{u=1}^{N} \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right)' E\left(e_{N,u}^2\right) D_N^{-1} e_1 \quad (1) \\
&+ e_1' D_N^{-1} \sum_{u=1}^{N} \sum_{\substack{u'=1 \\ u' \neq u}}^{N} \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \left(\sum_{v=u'}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right)' E\left(e_{N,u} e_{N,u'}\right) D_N^{-1} e_1 \quad (2)
\end{aligned}
$$

Now, following the proof of Proposition 1,

$$
\begin{aligned}
& \sum_{u=1}^{N} \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right)' \\
&= \lim_{N \to \infty} \sum_{u=1}^{N} \frac{1}{Nh} \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \left(\sum_{v'=u}^{N} \frac{1}{Nh} k\left(\frac{v'}{Nh}\right) X\left(\frac{v'}{Nh}\right) \right)' \left[\Delta V'\left(\frac{u}{N}\right) - 2\Delta_1 C\left(\frac{u}{N}, \frac{u}{N} - \frac{1}{N}\right) \right] \\
&= \sigma(0) \left(\int_0^\infty \left(\int_u^\infty k(v) X(v)\, dv \right) \left(\int_u^\infty k(v) X(v)\, dv \right)' du \right) \left(\lim_{N \to \infty} h \right)
\end{aligned}
$$

so

$$
(1) = O(h(N))
$$

and

$$
\begin{aligned}
& \sum_{u=1}^{N} \sum_{\substack{u'=1 \\ u' \neq u}}^{N} \left(\sum_{u=1}^{N} \frac{1}{Nh} \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \right) \left(\sum_{u'=1}^{N} \frac{1}{Nh} \left(\sum_{v=u'}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \right)' E\left(e_{N,u} e_{N,u'}\right) \\
&\leq \sup_{u \leq N} E\left(e_{N,u} e_{N,u'}\right) \sum_{u=1}^{N} \sum_{u'=1}^{N} \left(\sum_{u=1}^{N} \frac{1}{Nh} \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \right) \left(\sum_{u'=1}^{N} \frac{1}{Nh} \left(\sum_{v=u'}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \right)' \\
&= h(N)^2 O(1) \left(\sum_{u=1}^{N} \frac{1}{Nh} \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \right) \left(\sum_{u=1}^{N} \frac{1}{Nh} \left(\sum_{v=u}^{N} \frac{1}{Nh} k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \right)' = O\left(h(N)^2\right)
\end{aligned}
$$

Therefore,

$$
\frac{1}{h} \operatorname{var}\left(\alpha_N^+\right) \to \sigma(0) \left(\int_0^\infty \left(\int_u^\infty k(v) X(v)\, dv \right) \left(\int_u^\infty k(v) X(v)\, dv \right)' du \right) =: V_2
$$

since the second term is $O\left(h^2\right)$.

Asymptotic Distribution

I now consider the asymptotic distribution of $\frac{1}{\sqrt{h}}\left(\alpha_N^+ - y_+(0)\right)$:

$$
\begin{aligned}
\frac{1}{\sqrt{h}}\left(\alpha_N^+ - y_+(0)\right) &= \frac{1}{\sqrt{h}} e_1' D_N^{-1} \left(\sum_{u=1}^{N} \left(\frac{1}{Nh}\right) k\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right) \left(F\left(\frac{u}{N}\right) - F_+(0) \right) \right) \quad (1) \\
&+ \frac{1}{\sqrt{h}} e_1' D_N^{-1} \left[\sum_{u=1}^{N} \left(\sum_{v=u}^{N} \left(\frac{1}{Nh}\right) k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) e_{N,u} \right] \quad (2)
\end{aligned}
$$

Term (1), the bias, is $O\left(\sqrt{h}\right)$, and therefore goes to zero. Term (2) is a sum of mean-zero random variables:

$$(2) = \sum_{u=1}^{N} A_N(u) + o_p(1)$$

where

$$A_N(u) = \frac{1}{\sqrt{h}} \left(e_1' D_N^{-1} \sum_{v=u}^{N} \left(\frac{1}{Nh}\right) k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) e_{N,u}$$

Now, by Assumption 4, the increments $e_{N,u}$, and hence, $A_N(u)$ are associated. For associated random variables (Charles M. Newman, 1984), we have the following condition:

$$\left| E\left(\exp\left(it\sum_{v=1}^{N} A_N(v)\right)\right) - \prod_{v=1}^{N} E\left(\exp\left(itA_N(v)\right)\right) \right| \le t^2 \sum_{v=1}^{N} \sum_{\substack{v'=1 \\ v\neq v'}}^{N} \operatorname{cov}\left(A_N(v), A_N(v')\right)$$

and $\sum_{v=1}^{N} \sum_{\substack{v'=1 \\ v\neq v'}}^{N} \operatorname{cov}\left(A_N(v), A_N(v')\right)$ can be expressed as

$$\frac{1}{h} e_1' D_N^{-1} \sum_{u=1}^{N} \sum_{\substack{u'=1 \\ u'\neq u}}^{N} \left(\sum_{v=u}^{N} \left(\frac{1}{Nh}\right) k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right) \left(\sum_{v=u'}^{N} \left(\frac{1}{Nh}\right) k\left(\frac{v}{Nh}\right) X\left(\frac{v}{Nh}\right) \right)' E\left(e_{N,u} e_{N,u'}\right) D_N^{-1} e_1 = O\left(h(N)\right)$$

Since

$$\prod_{v=1}^{N} E\left(\exp\left(itA_N(v)\right)\right) \rightarrow_N \Phi\left(t, 0, V_2\right)$$

where $\Phi\left(z, \mu, \sigma^2\right)$ is the Gaussan distribution function with mean μ and variance σ^2, we have

$$\left| E\left(\exp\left(it\sum_{v=1}^{N} A_N(v)\right)\right) - \Phi\left(t, 0, \operatorname{var}\left(A_N\right)\right) \right| \le o(1) + O\left(h(N)\right) t^2 \rightarrow_N 0 \text{ for each } t.$$

Hence,

$$\frac{1}{\sqrt{h(N)}} \left(\alpha_N^+ - y_+(0)\right) \rightarrow^d N\left(0, V_2\right)$$

Feasible Estimation

The formula

$$V := \sigma(0)\, e_1' D^{-1} \left[\left(\int_0^\infty \left(\int_u^\infty k(v) X(v)\, dv \right) \left(\int_u^\infty k(v) X(v)\, dv \right)' du \right) \right] D^{-1} e_1$$

contains the unknown constant $\sigma(0)$ that must be estimated. I argue that the estimator:

$$\bar{V}_N = e_1' D_N^{-1} \left[\frac{1}{h} \sum_{u=1}^{N} k\left(\frac{u}{Nh}\right) X\left(\frac{u}{Nh}\right) y_{N,u}^2 \right]$$

is a feasible and consistent estimator of $\sigma(0)$ under Assumption 2'.
To prove this, we compute the moments of $\bar{V}_N$ and show that $E\left(\bar{V}_N\right) \rightarrow \bar{V}(0)$, while $E\left(\bar{V}_N^2\right) \rightarrow \left[E\left(\bar{V}_N\right)\right]^2$.
First, we see that $\bar{V}_N$ is asymptotically unbiased for $\sigma(0)$:

$$
\begin{aligned}
E\left(\bar{V}_N\right) &= e_1' D_N^{-1}\left[\sum_{u=1}^{N}\frac{1}{h}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)E\left[\left(y\left(\frac{u}{N}\right)-y\left(\frac{u-1}{N}\right)\right)^2\right]\right] \\
&= e_1' D_N^{-1}\left[\sum_{u=1}^{N}\frac{1}{h}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\left(F_{N,u}^2+E\left(e_{N,u}^2\right)\right)\right] \\
&= e_1' D_N^{-1}\left[\sum_{u=1}^{N}\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)E\left(Ne_{N,u}^2\right)\right]+O\left(\frac{1}{N}\right)
\end{aligned}
$$

Now, invoking Lemma 4, it is clear that $\lim_{N\to\infty} e_1' D_N^{-1}\left[\sum_{u=1}^{N}\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)E\left(Ne_{N,u}^2\right)\right] = E\left(Ne_{N,0}^2\right) = \sigma(0)$, since $\sigma(s)$ is continuous. Hence, $E\left(\bar{V}_N\right) \to \bar{V}(0)$.

To compute the variance of $\bar{V}_N$, note that

$$
N^2 E\left(y_{N,u}^2 y_{N,u'}^2\right) = F_{N,u}^2 F_{N,u'}^2 + E\left(e_{N,u}^2 e_{N,u}^2\right) = N^2 E\left(e_{N,u}^2 e_{N,u'}^2\right) + O\left(\frac{1}{N^2}\right)
$$

Therefore, we need to show only that

$$
\begin{aligned}
& e_1' D_N^{-1}\left[\sum_{u=1}^{N}\sum_{u'=1}^{N}\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(\frac{1}{Nh}k\left(\frac{u'}{Nh}\right)X\left(\frac{u'}{Nh}\right)\right)' N^2 E\left(e_{N,u}^2 e_{N,u'}^2\right)\right] D_N^{-1} e_1 \\
\to & \left(e_1' D_N^{-1}\left[\sum_{u=1}^{N}\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)E\left(Ne_{N,u}^2\right)\right]\right)^2
\end{aligned}
$$

Under the assumption

$$
\operatorname{cov}\left(Ne_{N,u}^2, Ne_{N,u'}^2\right) = O\left(\operatorname{cov}\left(\sqrt{N}e_{N,u}, \sqrt{N}e_{N,u'}\right)^2\right)
$$

(satisfied for a Gaussian process) we have

$$
E\left(Ne_{N,u}^2 Ne_{N,v}^2\right) = O\left(\operatorname{cov}\left(\sqrt{N}e_{N,u}, \sqrt{N}e_{N,u'}\right)^2\right) + E\left(Ne_{N,u}^2\right)E\left(Ne_{N,v}^2\right)
$$

Therefore,

$$
\begin{aligned}
& e_1' D_N^{-1}\left[\sum_{u=1}^{N}\sum_{u'=1}^{N}\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(\frac{1}{Nh}k\left(\frac{u'}{Nh}\right)X\left(\frac{u'}{Nh}\right)\right)' N^2 E\left(e_{N,u}^2 e_{N,u'}^2\right)\right] D_N^{-1} e_1 \\
& -\left(e_1' D_N^{-1}\left[\sum_{u=1}^{N}\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)E\left(Ne_{N,u}^2\right)\right]\right)^2 \\
= \; & e_1' D_N^{-1}\left[\sum_{u=1}^{N}\sum_{u'=1}^{N}\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(\frac{1}{Nh}k\left(\frac{u'}{Nh}\right)X\left(\frac{u'}{Nh}\right)\right)' O\left(\operatorname{cov}\left(\sqrt{N}e_{N,u}, \sqrt{N}e_{N,u'}\right)^2\right)\right] D_N^{-1} e_1 \\
= \; & e_1' D_N^{-1}\left[\sum_{u=1}^{N}\sum_{u'=1}^{N}\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)\left(\frac{1}{Nh}k\left(\frac{u'}{Nh}\right)X\left(\frac{u'}{Nh}\right)\right)' O\left(\frac{1}{N^2}\right)\right] D_N^{-1} e_1 \to 0.
\end{aligned}
$$

since $\operatorname{cov}\left(\sqrt{N}e_{N,u}, \sqrt{N}e_{N,u'}\right) = E\left(Ne_{N,u}e_{N,u'}\right) = O\left(\frac{1}{N}\right)$. Consistency is proved.

Covariance of α_N^+ and α_N^-

Under Assumption 2, it was obvious that the leading stochastic term of α_N^+ was composed of independent random variables, and that the covariance between α_N^+ and α_N^- is zero. However, with Assumption 2', the leading stochastic term of α_N^+ involves correlated random variables. Here, we check that the asymptotic covariance between α_N^+ and α_N^- remains zero.

The above equation is valid so long as $\text{cov}\left(\alpha_N^+, \alpha_N^-\right) = o\left(h\left(N\right)\right)$. To prove this, I extend Assumption 3', so that for $s > 0$ and $t < 0$, we have the following covariance structure:

$$E\left(e\left(s\right)e\left(t\right)\right) = C\left(s,t\right)$$

where $C\left(s,t\right)$ satisfies the hypotheses of Assumption 3'. Then, define

$$C_N^0\left(u\right) = \sum_{u=1}^{N}\left(\frac{1}{Nh}k\left(\frac{u}{Nh}\right)X\left(\frac{u}{Nh}\right)\right)$$

We have

$$\begin{aligned}
\text{cov}\left(\alpha_N^+, \alpha_N^-\right) &= E\left(Z_N^+ Z_N^-\right) = E\left(e_1' D_N^{-1}\sum_{u=1}^{N}\sum_{u'=1}^{N} C_N^0\left(u\right)C_N^0\left(u'\right)' y_{N,u}^+ y_{N,u'}^- D_N^{-1} e_1\right) \\
&= e_1' D_N^{-1}\sum_{u=1}^{N}\sum_{u'=1}^{N} C_N^0\left(u\right)C_N^0\left(u'\right)' E\left(e_{N,u}^+ e_{N,u'}^-\right) D_N^{-1} e_1 \quad (1') \\
&\quad + e_1' D_N^{-1}\sum_{u=1}^{N}\sum_{u'=1}^{N} C_N^0\left(u\right)C_N^0\left(u'\right)' F_{N,u}^+ F_{N,u'}^- D_N^{-1} e_1 \quad (2')
\end{aligned}$$

It is obvious that $(2') = O\left(h\left(N\right)^2\right)$, since $F_{N,u}^+ = O\left(1/N\right)$, and $(1') = O\left(h\left(N\right)^2\right)$, because it contains only covariance terms.

2.10.6 Implementation

We are ultimately interested in running regressions of the form

$$y_{i,b}\left(0\right) = \alpha_b + X_{i,b}\gamma + \varepsilon_{i,b}$$

Therefore, under the assumption that $v_+\left(0\right) = v_-\left(0\right) = v\left(0\right)$, the independent term goes into the fixed effect and does not need to be predicted. Hence, I can compute the variance of $y_{i,b}\left(0\right)$ as $V_{i,b} = \bar{V}_N$. I then compute the variance of γ as the appropriate submatrix of

$$\bar{V} = \frac{N}{N-K}\left(\hat{X}'W\hat{X}\right)^{-1}\left(\hat{X}'W\left(\text{diag}\left(\hat{\varepsilon}^2\right) + \bar{V}\right)W'\hat{X}\right)\left(\hat{X}'W\hat{X}\right)^{-1}$$

where K is the number of regressors (including fixed effects), diag$\left(\hat{\varepsilon}^2\right)$ is a diagonal matrix of the squared residuals, W is a weight matrix, $\hat{X}$ is the matrix of regressors including the fixed effects. and $\bar{V}$ is a diagonal matrix of the estimated variances.

Chapter 3

World Welfare is Rising: Estimation Using Nonparametric Bounds on Welfare Measures

3.1 Introduction

[1]There is a substantial literature on estimating the evolution of the global distribution of income over time to assess whether global poverty and inequality are rising or falling. An important strand of the literature argues that while inequality between countries treated as observations of equal weight is rising, inequality between all people on the globe is falling, as some of the fastest-growing countries (China and India) have initially been among the poorest. This claim has been advanced by, e.g. Schultz (1998), Bhalla (2002), Bourguignon and Morrisson (2002), and most recently Sala-i-Martin (2002a and b, 2006) and Chotikapanich et al. (2007). Similarly, Bhalla (2002), Chen and Ravallion [2001, 2010] and Sala-i-Martin (2002a and b, 2006) document that world poverty has been falling since 1990 (or 1970). There also is an alternative part of the literature (Dikhanov and Ward [2001], Milanovic [2002, 2005, 2012] that contends that global inequality, even if measured between individuals, has increased. The results derived in the literature are often cited to buttress or undermine the contention that the recent period of globalization has been good for the global poor, e.g. Bhalla (2002) or Milanovic (2005).

A common feature of this literature is that all its results are computed using grouped data on within-country inequality, typically obtained from a secondary dataset such as that of Deininger and Squire (1996) or its successor, the World Income Inequality Dataset (WIID). This has been done because data on income distributions, particularly for developing countries, is typically available only through tabulations, quintile shares, or Gini coefficients, with the microdata either not existing (as with very old surveys) or not being available to the public (as with many surveys administered by national statistical agencies, including those of critically important countries such as China).[2] While the World Bank has released some unit record data through its Living Standards

[1]This article (with minor differences) has been published in the January 2013 issue of the *Journal of Public Economics*. Elsevier grants permission for its authors to reproduce their articles in theses, as is explained on its Author Rights and Responsibilities page, http://www.elsevier.com/authors/author-rights-and-responsibilities

[2]Reddy and Minoiu (2007) comment on the paucity of unit record data: "The analysis of unit data may be

Measurement Surveys (LSMS), it is extremely difficult to find, not to mention exploit, a sufficient number and variety of surveys to obtain even one survey per country, let alone come close to the degree of coverage provided by a panel dataset like the WIID. The authors writing in the literature on estimating the world distribution of income implicitly or explicitly use parametric distributional assumptions to convert the grouped data into a distribution of income. Bourguignon and Morrisson (2002) and Milanovic (2002) assume that the distribution inside each quintile or decile[3] is egalitarian. Bhalla (2002) and Chen and Ravallion [2001, 2010] use parametrized Lorenz curves (for instance, Bhalla (2002) uses the World Bank's Simple Accounting Procedure). Dikhanov and Ward (2001) approximate the income distribution using a polynomial approximation, and Chotikapanich et. al. (2007) fit rich parametric distributions, such as the four-parameter generalized beta distribution, to income data.

However, there is no consensus on what is a good parametric assumption for the distribution of income. The literature on functional forms for income distributions is large and largely inconclusive, going back to Pareto (1897) for the Pareto distribution; Gibrat (1931), Kalecki (1945), Aitchitson and Brown (1957), and more recently Lopez and Servén (2006) for the lognormal distribution, Salem and Mount (1974) for the gamma distribution, Singh and Maddala (1976) for the Singh-Maddala distribution, and McDonald and Xu (1995) on the generalized beta family of distributions, nesting all the above. All of the above distributions are unimodal; Zhu (2005) has suggested that empirical income distributions may be multimodal, which opens the door to further candidate distributions.[4] Given the number of distributions considered as plausible candidates, it is hard to be particularly confident about any given parametric assumption.

Parametric assumptions have a deeper methodological problem: since they yield point estimates of income distributions, parametric assumptions force the researcher to reach a conclusion on the evolution of the distribution of income. Yet, it may be the case that the data are not fine enough to reach any conclusion; there may exist valid income distributions that generate the data, yet imply that the overall distribution of income has widened, and there may also exist equally valid distributions that also generate the data and imply the overall distribution has narrowed. A critical question is: can we know if this is ever the case? Yet more generally: if the functional forms that the literature has been assuming are wrong, it gives no guidance as to what alternative paths global poverty and inequality might have taken, and what paths they most certainly could not have taken. While there have been many proposed time paths of global poverty and inequality based on parametric assumptions, are there any paths of poverty and inequality measures that can be ruled out on the basis of the data we have?

Hence, it is interesting to ask whether we can dispense with parametric assumptions com-

prohibitive in terms of time and manpower, and since unit data may be unavailable for numerous country-years...unit data from nationally representative household surveys for many countries....are not publicly available.

[3] Hereafter; quintiles, deciles, and other partitions of the income distributions will be referred to as fractiles.

[4] An alternative approach is to use nonparametric estimators to obtain the world distribution of income strictly from the data. Sala-i-Martin (2002a and b, 2006) uses kernel density estimation on fractile means to obtain estimates for country income distributions, and integrates them to obtain the world income distribution. Such an approach avoids the critique of making arbitrary distributional assumptions that the parametric approach is subject to, and succeeds in obtaining an estimate for the world income distribution as a whole, rather than just of some of its statistics. However, the approach estimates income distributions consistently only as the grouping of the data becomes arbitrarily fine, while in practice, income distribution data is presented in only a few groups (5 or 10). Therefore, there is need for bounds on poverty and inequality measures of the world distribution of income that are valid for any underlying functional form of the individual country distributions.

pletely. In particular, to analyze a time series of the values of some functional of the world distribution of income (e.g. poverty or inequality), we do not want point estimates, but sharp upper and lower bounds on the value of this series at each point in time, so that there may exist income distributions compatible with the data such that these bounds are attained, but there exist no distributions compatible with the data that imply estimates for the series outside the bounds. Such a pair of sharp bounds would completely summarize what the data implies about the series in question: any series passing outside the bounds would be impossible, whereas any series contained in them would be conceivable.

Another feature of the literature on analyzing the world distribution of income is that it analyzes poverty and inequality separately, and typically reaches normative conclusions without a formal aggregation of the two in some theoretically justified manner. While different ways have been proposed to aggregate growth and inequality into a measure of welfare, no measure has been selected as definitive, and the papers looking at welfare measures are few (Pinkovskiy and Sala-i-Martin [2009], Atkinson and Brandolini [2010]). There has neither been an attempt to provide bounds for the evolution of such a measure based on weak and plausible assumptions. However, the need for such a measure in the normative analysis of the trends in the world distribution of income is crucial. It is clear that the assertion that inequality has risen does not imply that welfare has fallen for most reasonable notions of welfare; the rise in inequality may have accompanied a Pareto improvement. Similarly, an assertion that poverty has risen need not automatically imply that welfare has decreased, since such a connection would be valid only if the welfare function was concerned exclusively about the poor. Otherwise, rapid growth for a large number of people in other parts of the distribution could (for a suitable welfare function) offset the negative effects of a rise in poverty. Without a well-specified welfare function that is derived from clear axiomatic normative principles, we cannot rigorously weigh the differential benefits of poverty and inequality reduction.

The contributions of this paper are threefold. First, I derive sharp nonparametric bounds for the Atkinson welfare measure (which is commonly calculated and theoretically justified) in terms of the inequality statistics typically made available by statistical agencies: fractile shares and Gini coefficients. The formula for a tight upper bound to any inequality measure when fractile means and boundaries are known is well-known and is presented in the review by Cowell (2000). Cowell (2000) also reviews tightening of the bounds under the assumption that the density of income is monotonic decreasing in a given fractile, while Cowell (1991) provides tight upper bounds in cases in which the fractile boundaries are known, but the fractile means are unknown. However, to my knowledge, there has been no work on deriving the bounds for the Atkinson welfare index using the Gini coefficient, with or without fractile shares.[5] This problem is useful, since Gini coefficients are often reported by statistical agencies and used in the literature, and deriving bounds for the Atkinson welfare index based on the Gini coefficient allows the researcher to deduce something about the first measure from the second, both for empirical and theoretical purposes. The mathematical problem of deriving the bounds given the richest available data is nontrivial, but important to solve as taking advantage of all the available data substantially decreases the width of the bounds.

[5]The most related part of the literature derives tight upper bounds to the Gini coefficient based on fractile shares. Gastwirth (1972) provides a tight upper bound on the Gini coefficient given fractile means and boundaries, while Murray (1978) provides such a bound when fractile means are unknown, and Mehran (1975) calculates the bound when fractile boundaries are unknown.

This paper also belongs to the more general econometrics literature on nonparametric bounds, e.g. Manski (1995).

Second, I compute the welfare bounds for all possible countries and years using traditional GDP and inequality data from the literature: the Penn World Tables (PWT) and the World Income Inequality Database (WIID) maintained by UNU-WIDER. Notwithstanding that the ways to compute some of the bounds have been previously known, to my knowledge, there has not been any work systematically applying these bounds to the problem of calculating world welfare, and *a fortiori*, no previous work has used the sharp bounds that I derive. Under very conservative imputation assumptions for countries and years without inequality data, I aggregate up the country welfare bounds to obtain bounds for world welfare. My first and main conclusion is that between 1970 and 2006, world welfare has risen unambiguously. The lower bound on world welfare growth in the baseline specification implies that welfare rose by 88%, with the effect being equivalent to an increase in the income of every person in the world of over $2,600 in PPP-adjusted 2005 U.S. dollars. While the fact that world welfare measured in this way has increased may not be very surprising owing to the strong GDP growth enjoyed by the world over this period, the extent of growth is remarkable. Moreover, the sharper bounds that I derive allow a finer analysis of the consistency of welfare growth over the sample period: using these bounds, one can conclude that world welfare growth was always positive for any 10-year period under consideration, whereas under the bounds of Cowell (2000), such a conclusion could not be drawn. I subject my results to a battery of increasingly radical robustness checks that confirm that neither the finding that world welfare rose nor (to a lesser extent) the magnitude of this rise are sensitive to neither the necessary assumptions of my procedure nor the more substantive challenges to the validity of either the mean income or the survey data

A related result is that the traditional data used in calculating the world distribution of income cannot reject the hypothesis that world inequality has risen. For almost all variations in the methodology of computing the bounds, my estimates are compatible with world inequality rising, or world welfare growing slower than world GDP per capita. The lower bound estimate for the baseline scenario suggests that welfare rose by only 93% of what it would have potentially risen by under uniform GDP per capita growth. However, I can reject the hypothesis that rising inequality destroyed more than 50% of the welfare growth that would have obtained under uniform GDP per capita growth for most robustness checks for different methodologies of computing the bounds, and for many robustness checks I can reject the hypothesis that rising inequality destroyed more than 20% of potential welfare growth. In particular, no robustness check can yield a conclusion that inequality has unambiguously risen.

The paper is organized as follows: Section 2 presents the Atkinson measure of welfare and reviews its microfoundations. Section 3 presents the derivations of the nonparametric bounds. Section 4 discusses the data, the numerical implementation of the bounds, and the imputation assumptions made in order to construct bounds for world welfare. Section 5 presents the baseline results for world welfare and inequality, and discusses the gains from basing the bounds on more inequality statistics. Section 6 presents the robustness checks: 1) sampling error, 2) alternative imputation procedures, 3) alternative survey selections from the WIID dataset, 4) replacing the WIID dataset with the UTIP-UNIDO dataset pioneered by Galbraith and Kum (2005), 5) replacing the Penn World Tables GDP data by World Bank GDP data with different PPP, and 6) accounting for failure of survey coverage at the top. Section 7 concludes.

3.2 The Welfare Measure

An important aspect of the debate concerning the world distribution of income is the question of what are the appropriate metrics of poverty, inequality and welfare. In this paper, I will concentrate on a single family of welfare measures and its associated family of inequality measures, appealing to a well-known theoretical argument for the use of this family. This is the family of Atkinson equally-distributed income equivalents and inequality indices, introduced in Atkinson (1970). Atkinson treats the problem of assigning a welfare rating to an income distribution as equivalent to the problem of assigning utility ratings to lotteries. The Atkinson equally-distributed income (hereafter the welfare measure) is the certainty equivalent of the distribution of income treated as a lottery, whereas the Atkinson inequality measure is the risk premium divided by the mean income. If it is assumed that utility is CRRA with risk aversion γ, which is a standard assumption in most empirical work, and is empirically supported in e.g. Chiappori and Paiella (2006), the relevant welfare index becomes

$$W(\gamma) = \left(\int_0^{\infty} x^{1-\gamma} dF(x) \right)^{\frac{1}{1-\gamma}}$$

and the relevant inequality index is

$$A(\gamma) = 1 - \frac{W(\gamma)}{\mu}$$

where μ is the mean income.

In terms of choice over lotteries, $A(\gamma)$ is the relative risk premium of the income distribution.

Atkinson's index resolves the chief problem in the construction (or even conceptualization) of a welfare index: the need to make interpersonal comparisons. Instead of assuming that the evaluator has some social preferences that allow her to trade off some utilities against others, the evaluator treats the income distribution as a lottery out of which she must draw a prize, and values this distribution accordingly. The choice of the (perfectly selfish) evaluator is then the choice she would make behind the Rawlsian veil of ignorance (although the welfare index would be the Rawlsian SWF only for $\gamma \to \infty$).

It is obvious that for $\gamma \geq 1$, any income distribution with an atom at zero income produces maximum inequality (an Atkinson index of 1, or an equally-distributed income of zero), and it is also immediate that no allocation of fractile shares nor any value of the Gini coefficient can rule out the income distribution having an atom at zero income, so for all $\gamma \geq 1$, it is impossible to construct a lower bound for welfare. Moreover, as $\gamma \to 1$ from below, the lower bound continuously drifts towards zero. Hence, bounds based on the fractile shares and the Gini coefficient can only be constructed for $\gamma \in (0, 1)$. (In particular, using the Rawlsian SWF or attempting to test for first-order stochastic dominance cannot be done if one is to remain totally agnostic about the distribution as is done in this paper). I will use the central value in this interval, $\gamma = 0.5$, as the baseline for this paper. Statistics for the Atkinson inequality index for $\gamma = 0.5$ are routinely reported by developed countries (e.g. the United States Bureau of the Census reports Atkinson inequality indices for $\gamma = 0.25$, $\gamma = 0.5$ and $\gamma = 0.75$; the Luxembourg income study reports indices for $\gamma = 0.5$ and $\gamma = 1$), although I will show that the baseline result holds for γ as high as 0.9. I also perform a robustness check for values of γ higher than unity by assuming a lower bound for income equal to 1/5 of the lowest fractile mean. Under this assumption, my baseline results are valid for γ as high as about 1.5.

3.3 Analytical Derivation of Uniform Bounds

In this section, I will first list a few facts about Lorenz curves and about the formulation of the optimization problems that yield the bounds in Lorenz curve space. I will then solve the optimization problems when the Gini coefficient is specified. Finally, I will then describe the solutions of the problems with fractile shares only, which is known by the literature on bounding inequality measures, and is reviewed in Cowell (2000).

3.3.1 General Remarks

I first review a number of basic facts about Lorenz curves, which can be found in, e.g. Gastwirth (1972) or derived by inspection.

1. The Lorenz curve $L(p)$ of a nonnegative random variable distributed according to $F(x)$ is $L(p) = \frac{1}{\mu}\int_0^{F^{-1}(p)} x dF(x)$, where $p \in [0,1]$. This is an increasing and convex function, such that $L(0) = 0$ and $L(1) \leq 1$ (if a set of measure zero of the population holds a set of positive measure of income, then $L(1) < 1$).

2. Conversely, for any increasing and convex map $L(p)$ of $[0,1]$ into itself, there exists a distribution function $F(x)$ such that $L(p)$ is the Lorenz curve of the nonnegative random variable distributed according to $F(x)$.Hence, the set of Lorenz curves is

$$\mathfrak{L} = \{L \in C[0,1] : L \text{ is increasing, convex, } L(0) = 0,\ L(1) \leq 1\}$$

3. Any Lorenz curve is continuously differentiable at all but countably many points, and where it exists, $L'(p) = \frac{x}{\mu}$,where $x = F^{-1}(p)$.

4. The Gini coefficient of a distribution with Lorenz curve $L(p)$ is $G = 1 - 2\int_0^1 L(p)\,dp$.

5. The Atkinson welfare index of a distribution with Lorenz curve $L(p)$ is $W(\gamma) = \mu\left(\int_0^1 (L'(p))^{1-\gamma} dp\right)^{\frac{1}{1-\gamma}}$.

   Now, suppose we are given k fractile shares, or statements that individuals from the p_ith to the p_{i+1}st percentile (or in $[p_i, p_{i+1}]$) own fraction Q_i of the national income, with the cumulative share of national income owned by the lowest p_i earners being q_i. Suppose also that the mean of the income distribution is normalized to 1; then, the fractile boundaries, a_i^- and a_i^+ are defined as the (normalized) incomes of the p_ith and p_{i+1}st percentile of the income distribution respectively. The (normalized) mean income of fractile i is defined as $m_i = Q_i / (p_{i+1} - p_i)$.

6. By the definition of a Lorenz curve, an assignment of fractile shares equivalent to a set of constraints $L(p_i) = q_i$, $i = 1, ...k$, where p_i is the fraction of the population in or below fractile i, and q_i is the cumulative share of national income owned by this fraction of the population.

7. The statement that the boundaries of fractile i are $\left[a_i^-, a_i^+\right]$ is equivalent to the constraint $\lim_{p\downarrow p_i} L'(p_i) \geq a_i^-$ and $a_i^+ \geq \lim_{p\uparrow p_{i+1}} L'(p_i)$ (the inequalities are strict whenever there is no mass in the distribution of income around the p_ith or p_{i+1}st percentile).

Therefore, the main analytical problem of this paper can be formulated as follows:

$$\begin{aligned} \max_{L\in\mathfrak{L}} \text{ or } \min \int_0^1 \left(L'(p)\right)^{\alpha} dp \text{ st. } 1)\forall i &= 1,...,k,\ L(p_i) = q_i, \\ 2) \int_0^1 L(p)\,dp &= 0.5(1-G) =: \bar{G} \end{aligned} \tag{3.1}$$

where $\alpha \in (0,1)$
It will be useful to define the constraint set:
$\mathfrak{L}_c = \left\{ L \in \mathfrak{L} : \bar{G} = \int_0^1 L(p)\,dp \text{ and } \forall i = 1,..,k,\ L(p_i) = q_i \right\}$

3.3.2 Maxima

Maxima with Gini only

The optimization problem to maximize $W(1-\alpha)$ for given Gini is

$$\max_{L\in\mathfrak{L}} \int_0^1 \left(L'(p)\right)^{\alpha} dp \text{ st. } \bar{G} = \int_0^1 L(p)\,dp$$

Since the objective function is concave in L, the maximum is unique and is attained in the interior, so this is a standard problem in the calculus of variations, since $L(p)$ is a.e.-twice differentiable. To solve this problem, I form the Lagrangian

$$\mathcal{L} = \left(L'(p)^{\alpha}\right) - \lambda L(p)$$

and compute the Euler equation:

$$\frac{\partial \mathcal{L}}{\partial L} = \frac{d}{dp}\left(\frac{\partial \mathcal{L}}{\partial L'}\right) \Leftrightarrow \lambda = \alpha(1-\alpha)\left(L'(p)^{\alpha-2}\right) L''(p)$$

Proposition 5 *The solution to the optimization problem is given by*

$$L^*(p) = \frac{1-\alpha}{\lambda}\left(c_1 - \frac{\lambda}{\alpha}p\right)^{-\frac{\alpha}{1-\alpha}} - c_2$$

where the constants c_1, c_2 and the Lagrange multiplier λ can be calculated from the equations $L^*(0) = 0,\ L^*(1) = 1,\ \bar{G} = \int_0^1 L^*(p)\,dp$.[6]

Proof. *In text.* ■

Note that the solution is a convex function, and thus a valid Lorenz curve. Hence, the value of the optimum attained corresponds to a sharp lower bound.

[6] The implied CDF of this distribution is $F(x) = \frac{\alpha}{\lambda}\left(c_1 - \left(\frac{x}{\mu}\right)^{\sigma-1}\right)$ for $x \in \left[\mu(c_1)^{-\frac{1}{1-\alpha}}, \mu\left(c_1 - \frac{\lambda}{\alpha}\right)^{-\frac{1}{1-\alpha}}\right]$, $F(x) = 0$ for $x < \mu(c_1)^{-\frac{1}{1-\alpha}}$, and $F(x) = 1$ for $x > \mu\left(c_1 - \frac{\lambda}{\alpha}\right)^{-\frac{1}{1-\alpha}}$.

Maxima with Gini and fractile shares

The optimization problem to maximize $W(1-\alpha)$ for given Gini and fractile shares is

$$\max_{L\in\mathfrak{L}}\int_0^1 (L'(p))^{\alpha}\,dp \text{ st. } 1)\ \bar{G}=\int_0^1 L(p)\,dp \text{ and } 2)\ \forall i=1,..,k,\ L(p_i)=q_i \tag{3.2}$$

The Euler equation is the same as in problem 1, however, the constants c_1 and c_2 are now allowed to vary by the interval i. An upper bound could be obtained by calculating the values of $c_{1,i}$ and $c_{2,i}$ from the conditions $L(p_i)=q_i$, however, unless the sequence $\{c_{2,i}\}$ is monotone increasing, the resulting solution is not convex, and hence, does not belong to $\mathfrak{L}$. Hence, although it provides a greater lower bound than the solution when only the Gini constraint, or only the fractile shares are present, this bound is not sharp.

If in addition to the fractile shares, the fractile boundaries $\{a_i^-,a_i^+\}_{i=1}^k$ were given, a sharp upper bound could be obtained by solving the equivalent problem:

$$\max_{L\in\mathfrak{L}}\int_0^1 (L'(p))^{\alpha}\,dp \text{ st. } 1)\ \bar{G}=\int_0^1 L(p)\,dp,\ 2)\ \forall i=1,..,k,\ L(p_i)=q_i$$

$$3)\lim_{p\uparrow p_i}L'(p)\le a_i^- \text{ and } \lim_{p\downarrow p_{i+1}}L'(p)\ge a_i^+$$

for some $\{a_i^-,a_i^+\}_{i=1}^k$ where a_i^- is the lower bound of interval i, while a_i^+ is the upper bound of interval i. Now, define

$$R(p):=\max_{i=1,\dots,k}\left\{a_i^-(p-p_i)+q_i,a_i^+(p-p_{i+1})+q_{i+1}\right\}$$

and consider the modified optimization problem:

$$\max_{L\in\mathfrak{L}}\int_0^1 (L'(p))^{\alpha}\,dp \text{ st. } 1)\ \bar{G}=\int_0^1 L(p)\,dp,\ 2)\ \forall i=1,..,k,\ L(p_i)=q_i,\ 3)\ L(p)\ge R(p) \tag{3.3}$$

Since the optimal solution must be convex, we have

$$\begin{aligned}&\left(\lim_{p\uparrow p_i}L'(p)\le a_i^- \text{ and } \lim_{p\downarrow p_{i+1}}L'(p)\ge a_i^+\ \forall i=1,..,k,\right)\\ \Leftrightarrow\ &L(p)\ge\max_{i=1,\dots,k}\left\{a_i^-(p-p_i)+q_i,a_i^+(p-p_{i+1})+q_{i+1}\right\}=:R(p)\end{aligned}$$

so the two problems are equivalent.

By Kamien and Schwartz (1991), the solution to this latter problem is characterized as follows:

1. If $L(p)>R(p)$, then $L(p)=\frac{1-\alpha}{\lambda}\left(c_{1,i}-\frac{\lambda}{\alpha}p\right)^{-\frac{\alpha}{1-\alpha}}-c_{2,i}$ for some $c_{1,i}$ and $c_{2,i}$
2. If p is a "switching point" between $L(p)$ and $R(p)$ (so $\forall\varepsilon$, $\exists p_-,p_+\in N_\varepsilon(p)$ st. $L(p)>R(p)$ on $[p_-,p]$ and $L(p)=R(p)$ on $[p,p_+]$, or vice versa), then, $L(p)=R(p)$ and $L'(p)=R'(p)$.

Hence, the solution $L^*(p,\mathbf{a})$ taking the vector of fractile boundaries $\mathbf{a}:=\{a_i^-,a_i^+\}_{i=1}^k$ as given is characterized as follows:

Proposition 6 *$\forall i = 1,...,k$, $\exists$ p_i^-, p_i^+ : $p_i \leq p_i^- < p_i^+ \leq p_{i+1}$ st. $L^*(p,\mathbf{a}) = R(p)$ on $[p_i, p_i^-] \cup [p_i^+, p_{i+1}]$ and $L^*(p,\mathbf{a}) = \frac{1-\alpha}{\lambda}\left(c_{1,i} - \frac{\lambda}{\alpha}p\right)^{-\frac{\alpha}{1-\alpha}} - c_{2,i}$ on $[p_i^-, p_i^+]$, where λ solves $\bar{G} = \int_0^1 L^*(p,\mathbf{a})\,dp$, while $p_i^-, p_i^+, c_{1,i}$ and $c_{2,i}$ solve*

1. $\frac{1-\alpha}{\lambda}\left(c_{1,i} - \frac{\lambda}{\alpha}p_i^-\right)^{-\frac{\alpha}{1-\alpha}} - c_{2,i} = R\left(p_i^-\right) = a_i^-\left(p_i^- - p_i\right) + q_i$
2. $\frac{1-\alpha}{\lambda}\left(c_{1,i} - \frac{\lambda}{\alpha}p_i^+\right)^{-\frac{\alpha}{1-\alpha}} - c_{2,i} = R\left(p_i^+\right) = a_i^+\left(p_i^+ - p_{i+1}\right) + q_{i+1}$
3. $\left(c_{1,i} - \frac{\lambda}{\alpha}p_i^-\right)^{-\frac{1}{1-\alpha}} \geq a_i^-$ *with strict inequality iff* $p_i^- = p_i$*, and*
4. $\left(c_{1,i} - \frac{\lambda}{\alpha}p_i^+\right)^{-\frac{1}{1-\alpha}} \leq a_i^+$ *with strict inequality iff* $p_i^+ = p_{i+1}$.

Proof. *In text.* ■

The problem reduces to the finite-dimensional problem of optimization along the finite sequence **a**, which can be done with standard software. If the fractile boundaries are known, the formula can be used directly and no optimization is required.

Minima

The problem is

$$\inf_{L\in\mathfrak{L}} \int_0^1 \left(L'(p)\right)^\alpha dp \text{ st. } 1)\ \bar{G} = \int_0^1 L(p)\,dp,\ 2)\ \forall i = 1,..,k,\ L(p_i) = q_i \tag{3.4}$$

It can be shown that the infimum of this problem is attained by a Lorenz curve that is a linear spline with corners (possibly) at the points $\{p_i\}$ and with no more than one corner in any interval (p_i, p_{i+1}).

Proposition 7 *Consider the minimization problem (3.4). Then, the value of this problem is identical to the value of the following finite-dimensional problem:*

$$\min_{\{a_i^-,a_i^+\}_{i=1}^k} \left\{ \left(\sum\nolimits_{i=1}^k \left(\lambda_i \left(\frac{a_i^-}{\mu}\right)^\alpha + (1-\lambda_i)\left(\frac{a_i^+}{\mu}\right)^\alpha \right) \right)^{1/\alpha} \right\} \tag{3.5}$$

$$\text{st. } \forall i = 1,...,k,\ \lambda_i = \frac{a_i^+ - m_i}{a_i^+ - a_i^-},\ a_i^- \leq m_i \leq a_i^+ \leq a_{i+1}^- \text{ and } \bar{G} = \int_0^1 L(p,\mathbf{a})\,dp$$

where $L(.,\mathbf{a})$ is the Lorenz curve defined by the sequence $\{a_i^-, a_i^+\}_{i=1}^k$.

Proof. See appendix. ■

The proof relies on the following lemma:

Lemma 8 *(3-2) Lemma: Suppose that $L \in \mathfrak{L}_c$ is piecewise linear with finitely many corners. Then, there exist numbers $\{a_i^-, a_i^+\}_{i=1}^k$ such that $\forall i = 1,...,k$, $a_i^- \leq m_i \leq a_i^+ \leq a_{i+1}^-$, and the Lorenz curvc*

$$\hat{L} = \max_{i=1,...,k} \left\{ \max\left\{ a_i^-(p - p_i) + q_i, a_i^+(p - p_{i+1}) + q_{i+1} \right\} \right\}$$

satisfies $\hat{L} \in \mathfrak{L}_c$ and $\int_0^1 \left(\hat{L}'(p)\right)^\alpha dp \leq \int_0^1 \left(L'(p)\right)^\alpha dp$.

Proof. *See appendix.* ■

The 3-2 Lemma states that any piecewise linear Lorenz curve with finitely many kinks that are off the fractile constraints can be replaced with a modification with only $2k$ such kinks that satisfies the constraints in the minimization problem (3.4), and decreases the value of the program. Heuristically, the minimizing curve must have no more than one kink in each interval, or no more than $2k$ kinks overall, which makes the family of possible minimizing curves finite-dimensional, and allows them to be computed by standard numerical methods.If the fractile boundaries are known (or assumed), the formula can be used directly and no optimization is required.

In the special case that there are no fractile constraints, it is easy to derive an explicit formula for the maximum Atkinson given the Gini coefficient, as the optimal curve must have no more than one kink.

Proposition 9 *Consider the minimization problem (3.4) and omit the fractile constraints. Then, the maximum value of the Atkinson index is given by*

$$\max\left(G, 1-(1-G)^{\frac{1-\alpha}{\alpha}}\right)$$

Proof. *See appendix.* ∎

3.3.3 Results with Fractiles Only

Cowell (1977) proves that if there are k fractiles, each with known mean m_i and fractile boundaries $[a_i^-, a_i^+]$, then the maximum welfare is attained at intrafractile egalitarianism: the distribution is concentrated at the fractile means, and the value of the problem is given by $\left(\frac{1}{k}\sum_{i=1}^{k}\left(\frac{m_i}{\mu}\right)^{\alpha}\right)^{1/\alpha}$. The minimum welfare is attained by complete concentration on the fractile boundaries, or by the solution to the finite-dimensional optimization problem *(3.5)* omitting the Gini constraint. In particular, a crude (non-sharp) approximation to the minimizer of welfare with fractiles only can be computed in closed form by setting $a_i^- = m_{i-1}$ and $a_i^+ = m_{i+1}$ for each i.

Figures 3.1 and 3.2 presents plots of some welfare-maximizing and welfare-minimizing Lorenz curves (the quintile shares that they are based on are denoted by bold circles in the diagrams). It is clear that the welfare-minimizing curves are all piecewise linear, while the welfare-maximizing curves that involve the Gini are nonlinear. Note how the crude approximations to both the welfare-maximizing and welfare-minimizing curves are nonconvex.

3.4 Implementation

3.4.1 GDP Data: Penn World Table

The Penn World Table (hereafter PWT) is one of the most cited sources for purchasing-power-parity-adjusted GDP data. The latest edition (version 7.0) has nearly comprehensive coverage of 189 currently existing countries since 1970 to 2009. I reconstruct GDP for currently nonexisting countries (e.g. the Soviet Union, Czechoslovakia, East Germany) by applying the growth rates of Penn World Tables version 5.6 to the implied GDP for these countries in version 7.0. This procedure is discussed in detail in Pinkovskiy and Sala-i-Martin (2009).

A major controversy in the literature is whether estimates of GDP should come from national accounts or from household surveys. Ahluwalia et al. (1979) pioneered the combination of national accounts GDP and survey inequality data, which is the dominant approach today, and is used, e.g. by Bourguignon and Morrisson (2002), Sala-i-Martin (2002a and b, 2006) and Bhalla (2002). Proponents of using national accounts to estimate the mean of the distribution of income argue that survey means tend to understate mean income and sometimes yield implausible implications (see e.g. the discussion in Bhalla (2002)). Moreover, national accounts estimates of GDP, and the Penn World Table in particular, are extensively used in cross-country research on growth and development: in particular, the seminal works of Barro and Sala-i-Martin (1992,1995); Barro (1999); Acemoglu, Johnson and Robinson (2001, 2002, 2005); and Banerjee and Duflo (2004) all use Penn World Table GDP, sometimes in conjunction with the Deininger-Squire dataset on inequality, whereas no such paper to my knowledge uses survey means. Other papers, such as Milanovic (2002) and Anand and Segal (2008), strongly criticize the use of national accounts on the grounds that it is inconsistent to take the distribution of income from one source and the mean of income from another. A practical consideration in favor of national accounts is that national accounts estimates are calculated using common methodology for virtually all countries and years, whereas survey means tend to be available for far fewer countries only in select years, so further assumptions are required in order to use a series of survey means. Since the focus of this paper is to present the methodology of the uniform bounds and to observe their implications for widely used data on the distribution of income, I will use national accounts as my source of GDP, but I will conduct a robustness check using survey means. Moreover, in the robustness check correcting for nonresponse, I will attempt to control for possible mechanisms that lead national accounts and survey means to diverge.

Figures 3.4 and 3.5 present a brief summary of the data. First, we see that GDP growth in 1970-2006 has been extraordinary – GDP has nearly doubled.[7] Second, we see that between-country inequality – the value of the Atkinson inequality index with $\gamma = 0.5$ that would obtain if the income distribution in each country were egalitarian – fell significantly, with much of the fall taking place after 2000. These results are suggestive of the claim that world welfare has increased, but are not conclusive, since if within-country inequality has increased by a substantial amount, welfare could have actually fallen. In fact, these results could be perfectly consistent with a "nightmare scenario" of a global elite, tiny in number but evenly distributed across nations, capturing most of the gains to growth in the past several decades.

3.4.2 Inequality Data: The World Income Inequality Dataset

The World Income Inequality Dataset (WIID), maintained by UNU-WIDER, is a significantly improved and expanded version of the Deininger-Squire (DS) dataset pioneered in 1996. It is probably the most comprehensive, and the most cited source on income inequality around the world[8], presenting over 5,200 surveys for over 150 countries and 79 years. Over 75% of the surveys listed took place after 1970. All the survey data reported include an estimate of the Gini coefficient, over 2,700 surveys contain quintile shares, and over 2,000 contain decile shares. Moreover, for nearly all the surveys, the database records the coverage of the survey of different parts of the country in question, the income concept asked for in the survey (income or consumption, gross income or net, whether in-kind income is included), the conversion factor used to obtain inequality between persons from household-level data, and the statistical agency conducting the survey and the researchers

[7] Note that the sample period for this paper ends before the beginning of the global recession in 2008.
[8] The paper introducing the dataset, Deininger and Squire (1996), has 1,884 citations on GoogleScholar.

reporting its results.[9] However, the WIID, as well as the DS dataset from which it was constructed, errs on the side of comprehensiveness of coverage rather than quality of surveys. Atkinson and Brandolini (2001) criticize the DS dataset for including poorly conducted and methodologically unclear surveys alongside well-organized ones, and criticize much research using the dataset for disregarding the noncomparabilities of surveys with different income concepts, different equivalence scales, and different underlying populations.[10] Hence, an important problem for any researcher using the dataset is to provide a method of selecting which surveys to use that avoids these pitfalls.

Choice of Surveys

I subdivide all surveys in the WIID into groups, hereafter surveygroups, within which all surveys 1) describe the same country, and the same geographic, demographic and socioeconomic population within the country, 2) have the same income concept, 3) collect income data on the same unit, use the same unit of analysis, and use the same equivalence scale to convert between the two if they are distinct, and 4) have the same primary and secondary source. I identify 2240 such surveygroups in the WIID, which means that each contains on average a little over two surveys, but some have much wider coverage than others. I then, instead of selecting separate surveys, select entire surveygroups on a heuristic basis by weighting the following considerations in approximately the following lexicographical order:

1. I give preference to those surveygroups that provide decile shares over those that provide only quintile shares, and I give preference to surveygroups providing fractile shares over those that only provide Gini coefficients,

2. I attempt to ensure that the surveygroups cover the longest date range for each country, and be well-distributed over the sample period, preferring a few surveys in each decade to thorough coverage of some periods at the expense of others,

3. I attempt to ensure that the surveygroups selected for each country have the same or similar income concepts, the same or similar equivalence scales and geographical extent, and the same primary source.

4. I attempt to maintain homogeneity in the characteristics of surveygroups selected across countries, with surveys asking about disposable income and equivalizing on the basis of household per capita being preferred. However, there are gross exceptions to this (e.g. India offers quintile shares only for consumption surveys).

[9] There may be concern that the Gini coefficients listed in the WIID are estimated from the presented fractile shares. I have checked that no Gini coefficient that the WIID presents is given by either the minimum or the maximum value of the Gini coefficient that is theoretically compatible with the fractile shares (formulas are given in Mehran (1975)). The documentation to WIID mentions explicitly for a few surveys (121 out of 2240 survey groups) that the Gini was constructed from the fractile shares. In results not presented, I have recomputed the baseline estimates excluding these Gini data, and the results are indistinguishable from the baseline results reported in this paper.

[10] However, it appears that the WIID has drawn lessons from the critique of the DS dataset; in personal communication, Tony Atkinson noted that "WIDER did a great deal to clean the original DS database; the WIID database is much less subject to the kind of criticisms that [Atkinson and Brandolini (2001)] made." (Tony Atkinson, personal communication, December 2009).

5. In view of the fact that I will need to interpolate and extrapolate to estimate inequality for years for which I do not have survey data, I choose some surveys outside the sample period (i.e. before 1970).

I also maintain some region-specific conventions aside from these four general principles. In particular, I use the Luxembourg Income Study surveys for OECD countries unless long and detailed series are available in the WIID from the countries' own statistical bureaus. For Latin American and Caribbean countries, I almost invariably use the surveys provided by the Socio-Economic Database for Latin America and the Caribbean (SEDLAC), following the recommendation of WIID. For the populous East Asian countries of China, India and Indonesia, I use survey data provided by the national statistical bureaus. Finally, owing to a dearth of surveys for Africa, particularly from the beginning and middle of the sampling period, I suspend many of the homogeneity requirements and often use consumption surveygroups when these offer more extensive coverage than income surveygroups do. Overall, I choose surveygroups containing 1094 surveys, of which 1011 lie in the sample period.

For some countries and years, the unit records from household surveys are publically available. Chen and Ravallion (2010) use a database of over 700 surveys, many of which were made available to them as microdata, whereas others were provided only in grouped data form. Their PovCal website contains a description of all the surveys, including whether their unit records were available, and the parametric estimates of their underlying income distribution obtained using the Kakwani-Podder method. Furthermore, the Luxembourg Income Study provides microdata for many household income surveys in the OECD. For all countries and years for which microdata is available, I use either the published inequality statistics (Atkinson inequality indices) directly, or (in the case of the Chen-Ravallion data) I compute Atkinson inequality indices from the parametric estimates obtained by Chen and Ravallion on the basis of the microdata they used, assuming that these estimates are probably very close to the actual values of the Atkinson inequality indices in the microdata. Using microdata decreases the width of my bounds slightly but noticeably in the period 1995-2005 (to which most of the available microdata corresponds), and does not affect them for the preceding period.[11]

Breadth of Coverage

As is intuitive from Figures 3.4 and 3.5, inequality does not tend to vary much over short periods of time, especially when compared to variation in GDP, so interpolation (as opposed, possibly, to extrapolation) procedures to impute inequality measures for years without data should be relatively reliable. Hence, while one intuitive measure of the breadth of coverage is the percent of the world population in the given year who are covered by surveys, a potentially better measure is the percent of the world population who are either covered by a survey in that year, or whose inequality measures will be obtained by interpolation (rather than extrapolation). Hereafter, I define the *core* to be the set of individuals who are so covered. Figure 3.6 presents these measures. While the direct coverage measure is highly erratic (depending significantly on whether China and / or India are covered in

[11] For the OECD countries with Luxembourg Income Study data, the Atkinson inequality indices are available only for $\gamma = 0.5$ and $\gamma = 1$. Whenever I compute Atkinson indices with other parameters, I use the $\gamma = 1$ index from the LIS. This does not affect the qualitative conclusions reached.

a given year), and tends to be below 60%, the size of the core as a percent of the world population is remarkably continuous, and tends to be above 70% until about 1998, and around 90% for most of the 1980s and 1990s. Hence, at least until the 2000s, coverage using the WIID is rather good.

3.4.3 Numerical Implementation

I have implemented numerically all the bounds described above except for the sharp bound for the maximum welfare given both the Gini coefficient and the fractile shares. In its place, I am reporting the crude upper bound for welfare based on Ginis and fractiles, which, while superior to both the Gini and the fractile upper bounds taken individually, may only be attained by curves that are not convex. All the other bounds for the maximum are very easy to implement, as they only rely on finding the root of a monotonic function in one variable. The sharp bounds for the minimum welfare given fractile shares (with or without the Gini) require numerical optimization over a long vector of arguments, the sequence $\{a_i^-, a_i^+\}_{i=1}^k$.[12]

3.4.4 Assumptions for Interpolation and Extrapolation

The bounds I have computed given fractile shares and the Gini coefficient say absolutely nothing in theory about the behavior of inequality in countries and years for which we do not have data, so the only fully conservative bound for those country-years is the trivial bound $[0, 1]$. However, it is accepted in the area of inequality research that inequality tends to change very slowly and very continuously,[13] so in practice, inequality data in a given year should give a great deal of information about inequality in that country in nearby years. The average coefficient of variation of the Gini within a surveygroup is only 0.06, and it does not exceed 0.41 for any surveygroup.

A plausible and easy-to-implement interpolation assumption is that inequality in any given year for which data is missing is bounded above and below by the inequality in the closest preceding and following years (hereafter, closest available years) with data available. Then, the upper bound of inequality for that year is the maximum of the upper bounds of the inequality in the closest available years, while the lower bound is the minimum of the lower bounds. For country-years outside the core, this method is tantamount to horizontal extrapolation of the bounds, which is problematic as it may artificially truncate rising trends in inequality. Hence, I use a more conservative extrapolation procedure that interpolates the upper and lower bounds linearly if this would result in the bounds widening further apart, and horizontally otherwise. To avoid upper and lower bounds from reaching implausible values, I bound them by the maximum upper and minimum lower bounds obtained from the data within each World Bank region [14]. As it is very rare that inequality rises or falls at a linear rate for an extended interval of time, it is plausible that such extrapolation would account for the possible dynamics of inequality at the ends of the sample period.

[12] It is possible to implement this optimization straightforwardly using the Matlab program **fmincon** on a standard PC. While the solution does depend on the initial value chosen for the optimization, and while the program occasionally fails to converge, the variation in the result as a function of the initial value is extremely small (the bound on the inequality index varies by less than 1% of the maximum value of this index with $\gamma = 0.5$ for most observations). As a compromise between speed and accuracy, I run the program for each survey for no less than twenty randomly selected starting values, and stopping at the first subsequent time the program converges.

[13] See e.g. Bhalla (2002) or Galbraith and Kum (2005). The latter source considers changes of 5 Gini points or more per year to be "unlikely, except when they coincide with moments of major social upheaval."

[14] For a classification of countries into World Bank regions, see Sala-i-Martin (2006)

A more difficult problem is to impute inequality for countries with no survey data at all. Reasoning that countries may tend to be like other countries around them, and following Sala-i-Martin (2006) and Pinkovskiy and Sala-i-Martin (2009), I impute these inequality measures on the basis of the inequality of the other countries in their World Bank region. However, to be conservative, I impute the upper bound in every year to be the maximum upper bound observed in the data for all countries and years in that region, and the lower bound similarly. I will investigate more and less conservative methodologies for interpolation in the robustness checks.

3.5 Baseline Results

3.5.1 A Simple Test for a Rising or Falling Series

The results for each time series will take the form of upper and lower bounds, rather than point estimates. Under the assumptions for imputation, as well as under the assumption that the data is valid, and that sampling error can be ignored, these bounds contain the true value of the measure of interest with probability 1. Any path of the measure that is contained within the bounds is therefore consistent with the data, whereas any path that violates the bounds at any point is inconsistent with it.

There is a simple procedure for drawing conclusions as to whether a series increased or decreased between two dates. If the lower bound of the series at the earlier date exceeds the upper bound at the later date, the series fell for sure (with the caveats expressed in the previous paragraph). If, on the contrary, the *upper* bound of the series at the earlier date is exceeded by the lower bound of the series at the later date, the series rose for sure. However, neither of these statements may be true; in which case, it is impossible to draw conclusions on whether the series rose or fell between the two dates without further assumptions or data.

3.5.2 Example for a Single Country: Chinese Inequality

Before presenting my baseline results for inequality and welfare in the world as a whole, I present upper and lower bounds on the Atkinson inequality index for China in order to demonstrate the interaction of my bounding technique, interpolation and extrapolation on a single consistent set of surveys. Moreover, estimates for China are interesting in their own right because microdata from Chinese official income surveys is not released to the public. Figure 3.3 presents the series for the $\gamma = 0.5$ Atkinson inequality index for China. We see that inequality in China was between 0.07 and 0.24 in 1970 and 0.16 and 0.28 in 2006, which is consistent with Chinese inequality rising or falling over this time period (which involved a transition to capitalism and is believed to have witnessed a substantial rise in Chinese inequality). We also see that the interpolation and extrapolation procedures appear to be reasonable and to yield results that are not radically different from the observed values of the bounds.

3.5.3 Baseline results

All the baseline results and robustness checks are summarized in Table I. The table presents for all variations (except sampling error) 1) the minimum and maximum amounts by which Atkinson welfare (interpreted as the certainty equivalent of the world distribution of income) increased between 1970 and 2006, 2) the minimum and maximum percentage increases in Atkinson welfare since 1970, and 3) the minimum and maximum percentage increases in Atkinson welfare as a percentage of what they would have potentially been if all incomes grew at the same rate (uniform GDP growth). Thus, the lower bound in part 3) is informative as to how much less welfare growth there is because of the fact that growth in GDP per capita is distributed unequally. Note that we can reject the hypothesis that inequality rose if and only if the lower bound in part 3) is greater than 100%; welfare grew faster than did GDP because inequality shrank.

I present the time series of world welfare for implied risk aversion $\gamma = 0.5$ and $\gamma = 0.9$ in Figure 3.7 For both indices of risk aversion considered, world welfare rose between 1970 and 2006. For $\gamma = 0.5$, we can also reach the conclusion that world welfare rose between 1990 and 2006, and even between 2000 and 2006. These are very important findings, since they establish that for plausible levels of risk aversion, (and even for relatively high ones, such as $\gamma = 0.9$, for which fully nonparametric bounds may be expected to be difficult to construct), the only series consistent with the data imply that even accounting for its uneven distribution, growth was sufficiently high relative to any increase in inequality that overall welfare rose. This conclusion is also intuitive given the more primitive facts of the dataset we use: if per capita GDP grew by nearly a factor of two, and between-country inequality fell substantially, and within-country inequality as measured by the Gini varied very little, the only way that welfare could have fallen was if movements in the Gini coefficient and in the fractile shares were unrelated to movements in the Atkinson index. No less important is it to note by how much welfare rose: Table I indicates that for $\gamma = 0.5$, welfare rose by at least 88% between 1970 and 2006.

From the time series of inequality in Figure 3.8, I must remain agnostic about the direction of world inequality: it is impossible to tell whether inequality rose or fell without additional assumptions. While I can almost reject the hypothesis that inequality rose according to $\gamma = 0.5$ (the relative risk premium of the income distribution could have, at most, risen from 0.396 to 0.414), it is obvious by inspection that for $\gamma = 0.9$, the data is consistent with many possible rising or falling time paths of inequality. In particular, this finding indicates that the large drop in between-country inequality could have been more than overridden by a rise in within-country inequality. However, these bounds also display the relatively limited feasible variation in inequality. Table I shows that for the baseline specification, rising inequality could have eroded at most 7% of the welfare benefits of GDP growth, and for a risk aversion coefficient even as high as $\gamma = 0.9$, the largest possible inequality increase could have decreased the growth rate of welfare relative to uniform growth by at most 31%.

The benefit of using uniform bounds is that this failure to reject should not be interpreted as a "null result," but rather as a criticism of the (amount and presentation of the) data. It indicates that, at least without stronger assumptions on the form and evolution of inequality within countries, it is impossible to tell whether inequality rose or fell. If we wish to reach a conclusion, what is required is more surveys, more finely presented. In the robustness checks, I will show that the failure to reject comes largely from the paucity of information in the surveys (from the width of the bounds when inequality data is given) rather than from the conservatism of my imputation assumptions.

3.5.4 Gain from fine bounds

It is useful to see how much we gain by basing our bounds on additional data, and how much we gain by using sharp rather than loose bounds. Figure 3.9 presents welfare estimates for $\gamma = 0.5$ using four methods: 1) crude bounds based on fractiles, 2) sharp bounds based on fractiles, 3) sharp bounds based on the Gini, and 4) bounds (crude upper bound and sharp lower bound) based on both the Gini and fractiles. We see that 1) including the Gini coefficient as well as the fractile shares in estimating the bounds decreases the width of the intervals (more in the earlier than in the later part of the sample because there are substantially more surveys with unit records in the later part of the sample), 2) sharpening the fractile-based bounds does not appreciably decrease the width of the intervals, and 3) while the lower bound on welfare based on the Gini is very poor, the upper bound on welfare based on the Gini is quite good, and can be superior to the upper bound based on fractile shares (which, in practice, tend to be decile shares). In particular, the upper bounds on welfare are all extremely close together, which is consistent with the idea that there is little gain in further improving the upper bound, so our omission of the sharp upper bound with the Gini and fractile shares is not a large loss.

It is reasonable to ask whether there is anything that we gain from using finer bounds. From Figure 3.9 , we see that if we could base our bounds only on the Gini coefficient, we could not reject the hypothesis that welfare fell between 2000 and 2006 even for $\gamma = 0.5$. We can also deduce from Figures 3.8 and 3.9 that if we did not have a formula for calculating the upper bound on inequality given both the Gini and fractile shares, the data would be consistent with a large, rather than a trivial, rise in inequality for risk aversion coefficient $\gamma = 0.5$, and would have been consistent with a rise in inequality for risk aversion coefficients much lower than 0.5. However, at least within our baseline results, there are no other immediately obvious hypotheses that critically depend on the use of the finer bounds.

The power of the finer bounds can, however, be seen if we turn to the analysis of the fine structure of the welfare time series. It is of interest to ask what we can say about the rate of welfare growth over the period 1970-2006. Given upper and lower bounds for welfare, bounds for welfare growth can be easily constructed without losing sharpness by computing the upper bound of growth as the growth between the lower bound at date 1 and the upper bound at date 2, and vice versa for the lower bound of growth. It is easy to see from Figure 3.10 that the width of the bounds exceeds the typical 1-year growth rate, so it is useful to compute average growth rates over long periods, such as 10 years; however, this procedure prevents us from talking about growth trends at the ends of the sample period.[15]. Figure 3.10 shows bounds on the growth rate using fractiles only, and using fractiles together with the Gini coefficient for averaging periods of 10 years. It is obvious that using the finer bounds connotes an important improvement; we can reject the hypothesis that average annual growth rates in any 10-year period in the sample were negative using the fine bounds, but not using the fractile-based bounds. Moreover, the width of the bounds shrinks considerably when the bounds are finer, and we can make some nontrivial statements about the level of growth in different time periods with the fine bounds, such as the average annual growth rates being bounded away from zero. Unfortunately, we cannot make any statements about welfare growth accelerating or decelerating during the sample period, a question of obviously great interest.

3.5.5 Higher Atkinson Parameters

As mentioned in section 3.2, it is impossible to construct nontrivial bounds for Atkinson welfare indices with coefficient greater than unity because the lower bound is zero whenever a distribution

[15]This limitation may actually be appropriate in practice, since the growth dynamics at the ends of the sample period may be products of extrapolation. However, it is clear we can have too much of a good thing, as with 20-year average growth rates, we lose more than half of our 36-year long sample period.

of income with positive mass at zero is allowed. However, bounds for higher degree Atkinson indices can be constructed under the assumption of a minimum income. A flexible procedure for selecting such a minimum income is to assume that the minimum income is a fixed fraction of the lowest fractile mean, which allows poorer countries to have lower minimum incomes than richer countries. Figures 3.11 and 3.12 show plots for Atkinson welfare indices with $\gamma = 1.25$ and $\gamma = 1.5$ under the assumption that the minimum income is one-fifth of the mean income of the lowest fractile. Since the magnitudes of the higher-parameter Atkinson indices are much lower than the magnitude of the $\gamma = 0.5$ Atkinson index, the series are plotted at different scales, but the growth dynamics of the higher-parameter Atkinson indices are clear. Even for these higher values of the Atkinson parameter, welfare rises unambiguously, although we see from the figures and from Table I that the bounds are much wider and are compatible with much smaller rises in welfare. For $\gamma = 1.25$, rising inequality could have destroyed 31% of potential welfare growth, and for $\gamma = 1.5$, it could have eroded as much as 69%.

3.6 Robustness Checks

The formulae for the uniform bounds given fractile shares and the Gini coefficient are derived analytically, and hence need to be checked for robustness only to the relaxing the assumptions underlying them. The substantive assumptions underlying the baseline results presented in the previous sections are as follows: 1) the GDP data in the PWT and the inequality data in the WIID selected as described in fact do describe accurately the true GDP and inequality measures of the countries in question, and 2) the interpolation and extrapolation method assumed in section 4 is a good approximation for the actual behavior of the time series in question. These assumptions will be scrutinized in what follows.

3.6.1 Sampling error in the fractile shares and the Gini coefficient

The derivations in section 3 took the fractile shares and Gini coefficient to be known without error; in fact, these quantities are survey estimates that depend on the sample collected, so there may be a nonzero probability that the true values of the Atkinson index are not contained in the bounds constructed from the empirical estimates. The idea that sharp bounds based on empirical estimates may fail to contain the population value for which they have been constructed is explored in McDonald and Ransom (1981) and is a serious problem. In the context of the paper, there is reason to believe that this problem is small, since the WIID provides information on the sample sizes of most of the listed household surveys, and these sample sizes are very large, with median sample size equal to 23,900. However, since the surveys are nonrandom samples, and in particular, probably have high degrees of clustering, the variances of the resulting estimates are higher than the corresponding variances would have been had the surveys been simple random samples.

I perform a robustness check for sampling error by the following procedure:

1. I assume all underlying country distributions of income to be lognormal with inequality parameter implied by the Gini coefficient.

2. I draw 100 simple random samples from each survey with sample size equal to 1/10th of the listed sample size in order to control conservatively for the variance-inflating effect of nonrandom sampling procedures.

3. I compute the decile share-based bounds (using the crude closed-form version of the bound for the minimum welfare) for each draw.

4. I compute sampling-error-adjusted confidence bounds for welfare as the upper bound plus 2 standard deviations, and the lower bound minus 2 standard deviations.

5. I aggregate these bounds to obtain bounds for the entire world using the baseline imputation assumptions.

This procedure is very conservative, as sampling error is likely to be independent (or very weakly correlated) across surveys, so aggregating all the lower and upper bounds considers the very unlikely case that sampling error (as opposed to systematic error) consistently was in a downward (or upward) direction for *all* surveys in the dataset. Hence, the resulting confidence bounds contain the true value of world welfare in a given year with a probability far higher than 95%. The graph of the resulting bounds, along with the bounds based on the lognormal fractiles without sampling error, are presented in Figure 3.13. It is obvious that adding sampling error, even in a highly conservative fashion, does not substantially affect the bounds.

3.6.2 More conservative interpolation / extrapolation

One of the substantive assumptions that had to be made in the aggregation of country estimates to get the world welfare estimates concerned the imputation of bounds for country-years without any inequality data. Our baseline assumption is that the survey data gives us all the peaks and the troughs of the time series, so observations for the missing country-years should be contained between the outer envelope of the bounds of the closest available observations. An (extreme) alternative methodology would be to compute the highest upper bound and lowest lower bound observed in the data for the given country, and assume that inequality in this country never violates these bounds. Hence, we relax our assumption that all the peaks and troughs of the inequality series are observed to the assumption that we observe the highest peak and the lowest trough. Extrapolation is still performed linearly, so as to allow inequality to grow to values not observed in the sample.

I present the resulting bounds for welfare along with the baseline bounds for $\gamma = 0.5$ in Figure 3.14 and Table I. It is clear that we can reject the hypothesis that welfare did not grow in favor of the hypothesis that it grew for most periods of interest, and over the course of the sampling period. The bounds do widen, and we see that rising inequality could have destroyed as much as 41% of potential welfare growth (although the bounds are compatible with inequality falling as well).

One may argue that I fail to reject the hypothesis that world inequality rose because my interpolation scheme is too conservative: in particular, 1) the bounds for countries without surveys are too wide since they capture uncertainty in the level of inequality in the country as well as uncertainty coming from functional form, 2) the outer envelope interpolation is too cautious, since inequality tends to rise smoothly, 3) the linear extrapolation is too conservative as it inflates uncertainty due to functional form. In particular, Sala-i-Martin (2006) and Pinkovskiy and Sala-i-Martin (2009) impute inequality for countries with no survey data using regional average inequality, while Milanovic (2002) and Chen and Ravallion (2001) (implicitly) interpolate and extrapolate horizontally by using surveys from nearby years to stand in for surveys in years of interest. Therefore, Figure 3.20 considers what happens to the baseline inequality series ($\gamma = 0.5$) when these assumptions are relaxed. One modification replaces the bounds for countries without data by the average (rather than the envelope) of the bounds for countries in the same region with data, and the second modification also interpolates the bounds within the core (rather than taking their envelope) and uses horizontal extrapolation of the bounds rather than linear extrapolation. For the second modification, it is possible to barely reject the hypothesis that world inequality rose (the lower bound in 1970 is 0.4006 and the upper bound in 2006 is 0.3997), but this is entirely a result of using horizontal as opposed to linear extrapolation. Hence, in order to understand whether inequality has risen or fallen since 1970, it is necessary to collect more recent survey data (some of which has probably not been processed in the case of recent surveys) in finer categorizations.

3.6.3 Alternative inequality data: different procedures for choosing surveys

As noted by, e.g. Atkinson and Brandolini (2001), in using the DS database or the WIID, it is crucial to select comparable surveys so as to avoid comparing the inequality of conceptually different distributions. While it is difficult to write a formula that can combine the various considerations that go into determining which surveys to select, the methodology for selecting surveys that I have presented in section 4 can be justly criticized for being heuristic and difficult to replicate. Therefore, I provide two alternative methodologies; one that seeks to ensure comparability of the surveys selected within each country at the cost of a substantial loss of coverage, and another that attempts to control for the range of sampling and nonsampling error in the computation of the Gini coefficient at the cost of not being able to use fractile shares to reduce the width of the distribution-free bounds.

The first (hereafter homogeneous) methodology entails selecting the surveygroup with fractile shares with the largest number of surveys within the sample period for each country, and taking surveys for that country only from the selected surveygroup. Hence, all surveys for a given country must be identical along all dimensions that are held fixed within a surveygroup: source, underlying population, unit of analysis and equivalence scale, and income concept. However, this methodology does not attempt to ensure homogeneity across countries, and recognizes that while within-country trends in inequality will be measured using comparable data, the levels of inequality in different countries will not necessarily be comparable. (Trying to ensure homogeneity across countries by further excluding surveys from the WIID would do violence to the procedures, as either China, which has almost exclusively income surveys, or the Indian subcontinent, which has almost exclusively consumption surveys, would be excluded). From Figure 3.21, we see that this methodology drastically restricts coverage; only for the 1980s is more than 70% of the world covered even indirectly (in the core), and (not shown) the inequality series for China stops in 1992.

The second methodology (hereafter the extreme Ginis methodology) involves ignoring the differences between all surveys in terms of income concept and unit of analysis (but acknowledging the differences in terms of the underlying population), and for each country-year, taking as the final bounds the outer envelope of the bounds based only on the Gini coefficient for each Gini coefficient presented in the WIID for that country-year. The extreme Ginis methodology conjectures that all the income concepts and equivalence scales in the surveys are imperfectly implemented, but the range of resulting estimates captures the Gini coefficient that would result from an ideal implementation of a consistent income concept and equivalence scale. The average standard deviation of the Gini estimates is 0.042, which far exceeds the time standard deviation of the Gini coefficient within a given surveygroup across multiple years (whose mean and median are approximately 0.02). Hence, it is plausible that, given the wide range of the Gini estimates, this range contains the true value of the Gini. Obviously, this methodology expands the coverage of the surveys: Figure 3.22

shows that more than 90% of the world population are in the core until 2000, and more than 80% until 2003, while about 60% are directly covered by surveys.[16]

Figure 3.15 shows bounds for world welfare using the homogeneous methodology, while Figure 3.16 shows the nonparametric bounds for world welfare for $\gamma = 0.5$ for both the baseline estimates and the extreme Gini estimates. Since the extreme Ginis methodology can use only Gini-based bounds, I use the Gini-based baseline bounds for comparison. These estimates are remarkably close to the baseline estimates, and yield the same implications; for the homogeneous survey selection it is possible to conclude that world welfare rose for every decade for $\gamma = 0.5$, and we see from Table I that any rise in inequality could have destroyed no more than 17% of potential welfare growth. The fact that the extreme Gini estimates largely coincide with the baseline estimates is not unexpected when one considers the great width of the Gini bounds, which dwarfs most empirically plausible ranges of the Gini coefficient. [17]

3.6.4 Alternative GDP data: Treatment of Chinese GDP

In 2007, in the wake of concluding a series of price surveys in the developing world, the World Bank revised the prices it used in its purchasing-power-parity adjustments, which led to major changes in its GDP series in the World Development Indicators, in particular, the lowering of Chinese and Indian GDP by 40% and 35% respectively. This development has been reviewed in the popular press (The Economist: Nov. 29, 2007; Dec. 19, 2007). The revision has been criticized, in particular on the grounds that it considered prices in urban China only. Penn World Tables version 7 fully incorporates these PPP revisions, but, mindful of the controversy of the new Chinese PPP estimates, reports two estimates for China: one based exclusively on Chinese national income accounts (version 1) and the 2005 ICP price survey for PPP adjustment, and the other one including some further PPP adjustments to compensate for the potentially nonrepresentative geographical character of the Chinese price surveys in the 2005 ICP (version 2). In my analysis, I have used the version 2 China series from the Penn World Tables because it delivers more conservative results. Figure 3.17 shows bounds for world welfare using the version 1 series for Chinese GDP. We see from Table I that using the version 1 series actually strengthens my conclusion: world welfare rises by at least 101% from 1970 to 2006, and in particular, rises by at least 107% of uniform growth, suggesting that if we use the version 1 series, we could actually reject the hypothesis that inequality rose. However, such a rejection would not be robust to alternative extrapolation and interpolation methodologies, so I interpret this result cautiously.

3.6.5 Alternative GDP data: World Bank GDP

To check for robustness to the source of GDP more radically, I re-estimate world welfare and inequality measures using World Bank estimates of GDP from the World Development Indicators (hereafter WB). Figure 3.18 presents the WB welfare estimates for $\gamma = 0.5$ along the baseline results. The bounds are extremely close to each other. One may conjecture that the PWT GDP

[16] In results not reported, I also consider replacing the WIID survey data with data on Gini coefficients from Galbraith and Kum (1999). The results are essentially identical to my baseline Gini results.

[17] In fact, the extreme Gini bounds are sometimes narrower than the baseline bounds. This is because using Gini coefficients increases the number of country-years with surveys, thus replacing very conservative imputation procedures with much narrower nonparametric bounds for these country-years.

series might do a better job of describing GDP earlier in the sample period while the WB GDP should do a better job later in the sample period, which would mean that welfare rose by less than either set of bounds would imply separately. We see that even taking the outer envelope of the nonparametric bounds for PWT GDP and WB GDP, we can conclude that welfare rose over the sample period, and we can establish more restrictive hypotheses as well (e.g. welfare rose from 1990 to 2006). Hence the WB series does not substantially change our results.

3.6.6 Alternative GDP data: World Bank Survey Means

As discussed in Section 3.4.1, there is disagreement in the literature on whether to combine national accounts GDP with survey data on inequality, or to use the survey mean as a measure of the mean of the income distribution. In particular, Milanovic (2005) and Chen and Ravallion (2001) use survey means in their calculations of world poverty and inequality. In this paper, I have used national accounts data in order to 1) remain comparable to most of the literature on the evolution of the world distribution of income and to the growth literature, and 2) avoid problems relating to the unavailability of survey means for many countries and years, given that the coverage of national accounts is nearly universal. In this section, I investigate the robustness of my results to the use of survey means in place of national accounts GDP.

I use the data on survey means from the World Bank's poverty calculator, PovCalNet, which is the most complete and consistent panel of survey means that I am aware of. However, even this panel does not match the nearly complete coverage of the Penn World Tables. For 79 major countries (including China, India, Nigeria, Argentina, Mexico and the former Soviet Union) I can obtain survey mean data from 1990 to 2004, extrapolating and interpolating the survey means using the methods of section 4.4. I use the World Bank national accounts data to construct a comparison sample of the same 79 countries during the same time period. There are substantial differences between the national accounts and the survey means, which typically result in lower income and slower growth in the survey means than in the national accounts. For example, China's annual rate of growth is more than 1 percentage point smaller if computed using survey means than using national accounts. There are many explanations for these differences (Deaton 2005, 2010), such as intentional and unintentional survey misreporting, problems in monetizing in-kind income, and inappropriate national accounting.

Figure 3.19 presents the sharp upper and lower bounds for the welfare series computed for the 79-country composite using 1) the survey means, and 2) Penn World Table national accounts data for the period 1990-2004. The increase and the growth rate of welfare in this sample is much lower than for the baseline because the sample covers a much shorter period of time – in fact, the average annualized growth rate in the 79-country sample is 1.6% per year, while the average annualized growth rate in the baseline sample is 1.77% a year. It is clear that whether one chooses to use national accounts or survey means, world welfare rises unambiguously during this period, and, in fact, would rise unambiguously if we restricted our analysis to some subperiods of this data, such as 1990-2000. Table I shows the minimum absolute rise in welfare and the minimum growth rates for these welfare series. In particular, it is clear that in this subsample of countries, and even using survey means, we can actually reject the hypothesis that inequality rose within the subsample of countries because the lower bound of the ratio of welfare growth to per capita GDP growth exceeds 100%.

3.6.7 Accounting for nonresponse and nonrepresentativeness

A major problem with the survey data is that the people who respond to the surveys might systematically differ from people who do not. This concern is raised in Deaton (2005), who argues that falling response rates to many household surveys, and the large discrepancy between the national accounts means and the survey means, are problems of the first magnitude for the validity of the surveys. A particularly worrisome problem is that rich people in developing countries are systematically not covered by surveys (e.g. they live in gated communities to which surveyors have no access, or they openly lie to the surveyors for fear that their truthful answers might be given to the government). Atkinson, Piketty and Saez (2009) make the general case for the importance of inequality at the very top of the distribution in calculations of global inequality, and Banerjee and Piketty (2005) argue that failure of the Indian NSS to cover the top of the Indian income distribution may account for as much as 20%-40% of the much-documented growing gap between Indian national accounts and survey means in the NSS.

Given that the Atkinson welfare measure is decomposable, a simple method to generate sharp bounds given systematic nonresponse is to divide the population into two parts, respondents and nonrespondents, and combine the bounds for respondents with worst-case assumptions about nonrespondents. Specifically, suppose that the survey represents a fraction λ of the population, who have mean income μ_c. Note that λ is bounded below by the response rate, but may in fact be larger than the response rate if it is possible to adjust for nonresponse within the survey. Let $z := \mu_c/\mu$, and let $A_{LB}(\gamma)$ and $A_{UB}(\gamma)$ be the upper and lower bound for the Atkinson inequality index computed on the basis of the survey data without adjusting for nonrepresentativeness. Then, the sharp nonrepresentativeness-adjusted bounds for $A(\gamma)$ are as follows:

$$\text{lower bound: } \hat{A}_{LB}(1-\alpha) = 1 - \left[\lambda (z)^{\alpha} (1 - A_{LB}(1-\alpha))^{\alpha} + (1-\lambda)\left(\frac{1-\lambda z}{1-\lambda}\right)^{\alpha}\right]^{1/\alpha}$$

$$\text{upper bound: } \hat{A}_{UB}(1-\alpha) = 1 - \lambda^{1/\alpha} z (1 - A_{UB}(1-\alpha))$$

with the restriction that $\lambda z < 1$. This restriction always holds if $z < 1$. It is apparent that $\hat{A}_{LB}(1-\alpha)$ increases in (λ, z, α), and $\hat{A}_{UB}(1-\alpha)$ decreases in these variables, so as surveys become less representative (λ falls) and as survey mean income falls further below the national accounts GDP (z falls), the confidence intervals widen. Therefore, if we want to make the broader assumption that $z \geq \bar{z}$ and $\lambda \leq \bar{\lambda}$ for some $\bar{z}$ and $\bar{\lambda}$, we just compute the upper bound, and set the lower bound at $\hat{A}_{LB}(1-\alpha) = A_{LB}(1-\alpha)$.

The variable λ has the intuitive meaning of the response rate. The variable z is the ratio of the mean income of respondents to true mean income. If λ is equal to unity, and $z < 1$, then z should be interpreted as the fraction of national income accounted for by the survey because $1 - z$ is the fraction of income owned by the small number of unsurveyed super-rich individuals.

Unfortunately, the WIID does not contain data on the response rates to the surveys (either in the database itself or in the documentation), and reports mean survey incomes very sporadically (only one survey mean is ever reported for China). Therefore, for the purposes of this robustness check, I will assume common values λ and z for the entire world, and compute bounds on their basis. Such bounds will, in a sense, be more informative, as these bounds will be valid for all values of λ and z higher than the values chosen, whereas attempting to retrieve z from the WIID would introduce additional sources of error. I let λ take on values in $\{0.80, 0.90, 1\}$, following Korinek et. al. (2003), who give a reasonable range for nonresponse of 10%-30%. I let z take on values in $\{0.75, 0.875, 1\}$ following Banerjee and Piketty (2005), who argue that the top 1% of the Indian income distribution hold approximately 12% of Indian national income, and following Atkinson, Piketty and Saez (2009) who document that the top 1% of the US income distribution hold a share of national income approaching 25%.

First, suppose that nonresponse is negligible ($\lambda = 1$) and is coming from a small group of super-rich individuals who are not captured by the surveys. Figure 3.23 plots the welfare bounds for the baseline case, for $z = 0.875$ and for $z = 0.75$. For either value of z, we can be confident that welfare rose. We can barely fail to reject that welfare fell for the outer envelope of all the bounds presented. Thus, under the assumption that the discrepancy between survey mean income and national accounts income can be explained by part of the growth enriching a very small and very rich minority, with full response otherwise, we can conclude that welfare rose even if the fraction of national income held by this minority rose, for instance, from 0 to 25% over the sample period. However, the bounds are consistent with rising measured inequality and a rising share of income held by the super-rich substantially eroding welfare gains relative to uniform growth: Table I shows that if $z = 0.875$, the realized welfare growth may have been only 68% of per capita GDP growth, and if $z = 0.75$, the realized welfare growth may have been only 43% of per capita GDP growth. These are of course much more demanding sets of bounds because they try to not only capture the uncertainty in the population measured by surveys, but also attempt to account for factors changing the composition and size of this population.

We now consider the more general case when only λ of the population is covered by any household survey, and its mean income is z of the true mean income. This situation is very general; in particular, there are no assumptions at all on exactly how nonrandom the sampling is, and how large or small inequality is in the fraction of the population not surveyed. However, if survey means are low and nonresponse is high, the implication is that an increasingly large and economically significant part of the population is not being covered, which leads to wide nonparametric bounds and an inability to reach any conclusions without further assumptions. Figure 3.24 presents the baseline bounds as well as the bounds for $(\lambda, z) = (0.9, 0.875)$, $(\lambda, z) = (0.9, 0.75)$ and $(\lambda, z) = (0.8, 0.875)$. It is clear that we are confident for each set of the bounds that world welfare rose, but only barely so in some cases. In particular, setting the nonresponse rate to 30% ($\lambda = 0.7$), which is deemed possible by Korinek et al. (2003), would prevent us from rejecting the hypothesis that welfare fell. While we can be confident that welfare rose if the nonresponse rate is sufficiently small and if the mean income of respondents is sufficiently close to total mean income, we cannot be confident that welfare rose essentially for any higher nonresponse rates, which might nevertheless be plausible.

3.7 Conclusion

In this paper, I presented formulae for sharp, nonparametric bounds for typical measures of inequality and welfare that can be computed from standard summary statistics of income distributions that are routinely provided to the public. These bounds are valid independently of the functional form of the underlying distribution of income. Hence, these bounds illustrate exactly the extent of knowledge about the inequality measures in question that we gain from our data; they render moot any questions about appropriate assumptions for the form of within-country or within-fractile income distributions and they help focus the debate over whether we have seen improvement in living standards onto issues of the validity of the data from which they were computed.

Using national accounts GDP estimates and WIID survey data, the nonparametric bounds imply that for any series compatible with these data, welfare must have risen, and risen substantially, but it is not possible to conclude whether world inequality rose or fell. This claim is supported through sensitivity analysis over the exact use of the WIID data as well as over the methodology used to compute the national accounts GDP estimates. In particular, it appears reasonable to conclude that notwithstanding the incomparability problems of surveys in the WIID, and notwithstanding recent issues in adjusting national accounts for PPP, welfare has risen over the period 1970-2006, as well as over most shorter periods of interest within this time. It is also unlikely that sampling error or vagueness in the income concepts used in administering the surveys could change this result, so long as surveys are fully representative of whatever income concept they are measuring. For most of my robustness checks, I can rule out substantial rises in inequality, but if I use extremely conservative interpolation procedures, or if I attempt to account for the possible deterioration of survey representativeness of the population, substantial rises in inequality that destroy around 50% of potential welfare growth are compatible with the estimated bounds.

Hence, the major challenge that remains to the claim that welfare rose consistently over the period 1970-2006 is the concern that the WIID surveys are significantly nonrepresentative of the underlying population because of selective nonresponse. While under some assumptions, this concern does not overturn our conclusion, for some possible estimates of the extent of nonresponse, we obtain that further assumptions are needed to conclude anything about the path of welfare during the sample period. It is therefore of great importance to accurately gauge the reliability of the survey data we are using in order to reach conclusions about what happened to welfare in recent times.

3.8 Proofs of Various Propositions

3.8.1 Proof of Lemma 1 (3-2 Lemma):

Suppose that the increasing and convex curve $L(p)$ on $[0, p^*]$, where $p^* > 1$, is defined by three line segments, where the first segment is defined by $q = mp$, and the third segment is defined by $q = zp - (z-m)$, where $0 \leq m < z \leq \infty$. Let the second segment lie between the point $(\bar{p}, m\bar{p})$ on the first line and $\left(1 + \frac{\bar{q}-m}{z}, \bar{q}\right)$ on the third line, so its equation is $M(p) := q = m\bar{p} + \frac{z(\bar{q}-m\bar{p})}{(z-m)+(\bar{q}-z\bar{p})}(p - \bar{p})$. Therefore, $L(p) = \max\left(mp, m\bar{p} + \frac{z(\bar{q}-m\bar{p})}{(z-m)+(\bar{q}-z\bar{p})}(p-\bar{p}), zp - (z-m)\right)$ on $[0, p^*]$. Let the value of the third line segment at p^* be defined as rp^*.

Then, the Gini constraint is

$$\int_0^{p^*} L(p)\, dp = K_1 + K_2\left[(\bar{q}-m)(1-\bar{p})\right] = \bar{G}$$

for some K_1 and K_2 that are constants in $\bar{p}$ and $\bar{q}$, so the Gini constraint is equivalent to $\bar{q} = \frac{S}{1-\bar{p}} + m$, for some $S \in [0, \infty]$. Note that $\bar{p}$ is then constrained to lie in the interval $[0, \tilde{p}]$ for some $\tilde{p} < 1$, since $\bar{q} \leq rp^*$

The objective is given by

$$O = \int_0^{p^*} \left(L'(p)\right)^\alpha dp = \bar{p}m^\alpha + z^{-(1-\alpha)}(\bar{q} - m\bar{p})^\alpha \left(z(1-\bar{p}) + \bar{q} - m\right)^{1-\alpha} + z^\alpha\left(K - \frac{\bar{q}-m}{z}\right)$$

.

If and only if it can be shown that O is minimized by some $\bar{p} \in \{0, \tilde{p}\}$, or that the problem of minimizing O in $\bar{p}$ subject to the Gini constraint yields a corner solution, then the lemma is proved; since then the line $M(p)$ dominates either the first or the third line on $[0, p^*]$. Hence, the lemma is equivalent to the problem

$$\min_{p \in [\tilde{p}, 1]} \left\{(1-p)\, m^\alpha + z^{-(1-\alpha)}\left(\frac{S}{p}\right)\left[\left(1 + \frac{z}{S}p^2\right)^\alpha \left(1 + \frac{m}{S}p^2\right)^{1-\alpha} - 1\right]\right\}$$

having a corner solution (where we replace $p = 1 - \bar{p}$). Now, let $w = p^2/S \in (0, \infty)$. Then, we obtain.

$D_p = z^\alpha\left[\left(\frac{1+mw}{1+zw}\right)^\alpha \left(2\left((1-\alpha) + \alpha\frac{m}{z}\frac{1+zw}{1+mw}\right) - 1\right) + \frac{1}{zw}\left(1 - \left(\frac{1+mw}{1+zw}\right)^\alpha\right) - \left(\frac{m}{z}\right)^\alpha\right]$, and

$D_p^2 = -\frac{2}{p}\left[D_p + z^\alpha\left(\left(\frac{m}{z}\right)^\alpha - \left(\frac{1+mw}{1+zw}\right)^\alpha \left(1 - \frac{1}{z}\left(\frac{\alpha(z-m)}{1+mw}\right)\left(1 + \frac{2(1-\alpha)(z-m)w}{(1+mw)(1+zw)}\right)\right)\right)\right]$

A necessary and sufficient condition for a corner solution to the optimization problem is that $D_p = 0 \Rightarrow D_p^2 < 0$, or that

$$\begin{aligned}\left(\frac{m}{z}\right)^\alpha &= \left(\frac{1+mw}{1+zw}\right)^\alpha \left(2\left((1-\alpha) + \alpha\frac{m}{z}\frac{1+zw}{1+mw}\right) - 1\right) + \frac{1}{zw}\left(1 - \left(\frac{1+mw}{1+zw}\right)^\alpha\right) \\ &\Rightarrow \left(\frac{m}{z}\right)^\alpha \geq \left(\frac{1+mw}{1+zw}\right)^\alpha \left(1 - \frac{1}{z}\left(\frac{\alpha(z-m)}{1+mw}\right)\left(1 + \frac{2(1-\alpha)(z-m)w}{(1+mw)(1+zw)}\right)\right)\end{aligned}$$

which is trivially true.

Hence, the minimum of the problem is achieved at the boundary, and the optimizing segment of the curve has only one interior corner rather than two. It is obvious by induction that for any piecewise linear curve with z corners between two consecutive constrained points, there exists another piecewise linear curve that has only one corner between these points, satisfies the constraints of the original curve, and attains a weakly smaller value of the program.

In particular, for any piecewise linear Lorenz curve L with finitely many kinks, there exists a sequence $\{a_i^-, a_i^+\}_{i=1}^k$ such that $a_i^- \leq m_i \leq a_i^+ \leq a_{i+1}^-$ $\forall i = 1, ..., k$, and the Lorenz curve given by

$$\hat{L} = \max_{i=1,\ldots,k} \{\max \{a_i^- (p - p_i) + q_i, a_i^+ (p - p_{i+1}) + q_{i+1}\}\}$$

satisfies the constraints and attains a weakly smaller value of the objective than does the curve L.

3.8.2 Proof of Proposition 2:

The problem is

$$\inf_{L \in \mathfrak{L}} \int_0^1 (L'(p))^\alpha \, dp \text{ st. 1) } \bar{G} = \int_0^1 L(p) \, dp, \text{ 2) } \forall i = 1, .., k, \; L(p_i) = q_i$$

Since the functional $\int_0^1 (L'(p))^\alpha dp$ is bounded below by zero, it must be the case that $S := \inf_{L \in \mathfrak{L}} \int_0^1 (L'(p))^\alpha dp \in \mathbb{R}$. Moreover, there must be a sequence $\{L_i\} \in \mathfrak{L}_c$ such that $\lim_{i \to \infty} \int_0^1 (L_i'(p))^\alpha dp = S$. Now, since the function $L_i'(p)$ is Riemann integrable, it must be the case that for any ε, and for any i, there exists $\tilde{L}_i \in \mathfrak{L}_c$ such that $\tilde{L}_i$ is piecewise linear with finitely many corners, and $\left| \int_0^1 (L_i'(p))^\alpha dp - \int_0^1 \left(\tilde{L}_i'(p)\right)^\alpha dp \right| \leq \varepsilon$.[18] Hence, let $\{\varepsilon_i\} \in \mathbb{R}$ be a sequence such that $\lim_{i \to \infty} \varepsilon_i = 0$, and let $\left\{\tilde{L}_i\right\}$ be a sequence of piecewise linear functions with finitely many corners such that $\left| \int_0^1 (L_i'(p))^\alpha dp - \int_0^1 \left(\tilde{L}_i'(p)\right)^\alpha dp \right| \leq \varepsilon_i$. Then, $\lim_{i \to \infty} \int_0^1 \left(\tilde{L}_i'(p)\right)^\alpha dp = S$.

Now, by Lemma 1, for every piecewise linear Lorenz curve $\tilde{L}_i \in \mathfrak{L}_c$, there exists a piecewise linear Lorenz curve $\hat{L}(;\mathbf{a}_i) \in \mathfrak{L}_c$ given by

$$\hat{L}(p; \mathbf{a}_i) = \max_{s=1,\ldots,k} \left\{\max \left\{a_{i,s}^- (p - p_s) + q_s, a_{i,s}^+ (p - p_{s+1}) + q_{s+1}\right\}\right\}$$

for some $\mathbf{a}_i = \left\{a_{i,s}^-, a_{i,s}^+\right\}_{s=1}^k$ such that $a_{i,s}^- \leq m_s \leq a_{i,s}^+ \leq a_{i,s+1}^-$, such that $\int_0^1 \left(\hat{L}'(p; \mathbf{a}_i)\right)^\alpha dp \leq \int_0^1 \left(\tilde{L}_i'(p)\right)^\alpha dp$. Hence, $\lim_{i \to \infty} \int_0^1 \left(\hat{L}'(p; \mathbf{a}_i)\right)^\alpha dp \leq S$, and by definition of infimum, $\lim_{i \to \infty} \int_0^1 \left(\hat{L}'(p; \mathbf{a}_i)\right)^\alpha dp = S$. Now, let the set $\mathbf{A}$ be the set of all $\mathbf{a}_i$ satisfying the restriction $a_{i,s}^- \leq m_s \leq a_{i,s}^+ \leq a_{i,s+1}^-$ $\forall i = 1, ..., k$, and note that this is a closed subset of the compact set $\bar{\mathbb{R}}_+^{2k}$, and is therefore compact. Hence, the sequence $\{\mathbf{a}_i\}_{i=1}^\infty$ has a convergent subsequence, $\left\{\mathbf{a}_{i(k)}\right\}_{k=1}^\infty$, which converges to a limit $\mathbf{a}$. Finally, observe that by the definition of $\hat{L}(p; \mathbf{a}_i)$, the integral $\int_0^1 \left(\hat{L}'(p; \mathbf{a}_i)\right)^\alpha dp$ is continuous in $\mathbf{a}_i$, so

$S = \lim_{i \to \infty} \int_0^1 \left(\hat{L}'(p; \mathbf{a}_i)\right)^\alpha dp = \lim_{k \to \infty} \int_0^1 \left(\hat{L}'(p; \mathbf{a}_{i(k)})\right)^\alpha dp = \int_0^1 \left(\hat{L}'(p; \mathbf{a})\right)^\alpha dp$

and $L(p; \mathbf{a})$ is a Lorenz curve that attains the infimum value S. Since $L(p; \mathbf{a})$ is defined by $2k$ parameters, its coefficients $\{a_i^-, a_i^+\}_{i=1}^k$ can be solved for using standard numerical methods.

[18] I am very grateful to Paolo Siconolfi for help with this part of the proof.

3.8.3 Proof of Proposition 3:

The 3-2 Lemma implies that if there is only a Gini constraint, the Atkinson is maximized by a Lorenz curve $L(p)$ with only one interior corner (p, q).

The Gini of this Lorenz curve is given by $G = p - q$, and the Atkinson is given by $1 - q^{\alpha} p^{1-\alpha} - (1-q)^{\alpha}(1-p)^{1-\alpha}$, so the parameters p and q of the optimal curve are given by

$$\bar{p} = \arg\min_{p \in [G,1]} \left\{ (p-G)^{\alpha} p^{1-\alpha} + (1-p+G)^{\alpha}(1-p)^{1-\alpha} \right\}$$

and $\bar{q} = p - G$. The second derivative of the minimand is given by

$$D_p^2 = -\alpha(1-\alpha)\,\bar{q}^{\alpha-1}\bar{p}^{-\alpha}\left[\left(\frac{\bar{q}}{\bar{p}} + \frac{\bar{p}}{\bar{q}} - 2\right) + \left(\frac{1-\bar{q}}{1-\bar{p}} + \frac{1-\bar{p}}{1-\bar{q}} - 2\right)\right]$$

which is negative, so any minimum must be a corner solution, and the maximized Atkinson is given by $\max\left(G, 1-(1-G)^{\frac{1-\alpha}{\alpha}}\right)$.

3.9 Table

Table I: Bounds on Increase and Growth of World Welfare						
	Diff. Lower Bound	Diff. Upper Bound	Growth Lower Bound %	Growth Upper Bound %	Rel. Growth, Lower Bound %	Rel. Growth, Upper Bound %
3.7 Baseline	2626	3361	88	127	93	134
3.7 Gamma=0.9	1257	3234	65	330	69	350
3.9 Gini Bounds Only	1949	3937	65	190	69	201
3.9 Fractile Bounds, Sharp	2501	3579	83	147	88	156
3.9 Fractile Bounds, Crude	2484	3623	82	152	87	161
3.11 Baseline, lowest income at 0.2 of lowest fractile mean	2610	3331	87	124	92	131
3.11 Gamma=1.25, lowest income at 0.2 of lowest fractile mean	1052	2035	67	167	71	176
3.12 Gamma=1.5, lowest income at 0.2 of lowest fractile mean	382	1585	29	172	31	182
3.14 Extreme Interpolation	1731	4011	56	167	59	176
3.15 Homogeneous Survey Choice	2365	3541	78	132	83	140
3.16 Maximum Ginis	1510	3760	50	190	53	201
3.18 Alternative PWT Series for China	2804	3507	101	142	107	150
3.18 World Bank GDP	2548	3222	90	128	96	136
3.19 Survey Means	187	305	26	47	107	193
3.19 PWT GDP for Survey Means sample	780	1171	35	58	143	238
3.23 Super-Rich have <12.5 percent	1910	3695	64	160	68	169
3.23 Super-Rich have <25 percent	1214	4024	40	203	43	215
3.24 <10 percent Nonresponse, >87.5% Mean Income Ratio	984	4133	33	221	35	234
3.24 <20 percent Nonresponse, >75% Mean Income Ratio	154	4524	5	306	5	324
3.24 <10 percent Nonresponse, >87.5% Mean Income Ratio	419	4399	14	275	14	291

Note: Table I summarizes the results from all the graphs. The number for every row indicates the graph from which the relevant bounds are taken. The first two columns present bounds on the absolute increase of world welfare (the certainty equivalent of the income distribution) in dollars. The next two columns present bounds on the aggregate growth rate of world welfare in percent. The last two columns present bounds on the ratio of the aggregate growth rate of world welfare to the growth rate of world GDP per capita in percent (so 100 percent would correspond to welfare growing by as much as GDP per capita, or uniform growth).

The variations are:

Row 3.7 Baseline: PWT 7 GDP, WIID surveys selected by the procedure described in section 4. Microdata used where available as described in section 4.

For country-years with no survey data: interpolate the bounds as the outer envelope of the bounds of the adjacent years with survey data;

extrapolate the bounds as the outer envelope of horizontal and linear extrapolation of each bound;

impute bounds for countries with one or no surveys as the outer envelope of the bounds for all countries in the given region at any time in the sample period.

Bounds based on Gini coefficient and fractile shares for all country-years with survey data. Atkinson parameter γ is 0.5

Row 3.7 γ=0.9: Same as Baseline, but γ=0.9

Row 3.9 Gini Bounds Only: Same as Baseline, but bounds based on only Gini coefficient.

Row 3.9 Fractile Bounds, Sharp: Same as Baseline, but bounds based on only fractile shares (exact computation)

Row 3.9 Fractile Bounds, Crude: Same as Baseline, but bounds based on only fractile shares (approximate computation)

Row 3.11 Baseline, lowest income at 0.2 of lowest fractile mean: Same as Baseline, but bounds computed under assumption that the lowest income is 0.2 of the lowest fractile mean.

Row 3.11 γ=1.25, lowest income at 0.2 of lowest fractile mean: Same as Baseline, but bounds computed under assumption that the lowest income is 0.2 of the lowest fractile mean.

Row 3.12 γ=1.5, lowest income at 0.2 of lowest fractile mean: Same as Baseline, but bounds computed under assumption that the lowest income is 0.2 of the lowest fractile mean.

Row 3.14 Extreme Interpolation: Same as Baseline, except for country-years with no survey data, impute the outer envelope of the bounds over the country over the sample period (if country has more than one survey), or impute the outer envelope of the bounds over the entire region over the sample period (if country has one or no surveys).

Row 3.15 Homogeneous Survey Choice: Same as Baseline, except choose longest survey series from the same source and with same equivalization and income concept for each country from WIID.

Row 3.16 Maximum Ginis: Same as Baseline, except for each country-year with any WIID survey data on a national scale, use envelope of all Gini-based bounds implied by WIID surveys in country-year.

Row 3.17 Alternative China GDP: Same as Baseline, except use the PWT 7 version 1 GDP series, which bases Chinese GDP exclusively on the 2005 PPP revision without further adjustments.

Row 3.18 World Bank GDP: Same as Baseline, except use GDP data from World Development Indicators, 2008.

Row 3.19 Survey Means: Same as Baseline, except use survey means from the WIID instead of PWT GDP for all years in which they are available.

Row 3.19 PWT GDP for Survey Means sample: Same as Baseline, except restrict to the sample of countries with survey mean data.

Row 3.23 Super-Rich have <12.5 percent: Assume WIID statistics exclude a set of people of measure zero who own at most 12.5% of national income respectively.

Row 3.23 Super-Rich have <25 percent: Assume WIID statistics exclude a set of people of measure zero who own at most 25% of national income respectively.

Row 3.24 <10 percent Nonresponse, >87.5% Mean Income Ratio: Assume WIID statistics exclude nonrespondents, with respondents' mean income at least 87.5% of the overall mean income but the distribution of this income among the nonrespondents is arbitrary.

Row 3.24 <20 percent Nonresponse, >87.5% Mean Income Ratio: Assume WIID statistics exclude nonrespondents, with respondents' mean income at least 87.5% of the overall mean income but the distribution of this income among the nonrespondents is arbitrary.

Row 3.24 <10 percent Nonresponse, >75% Mean Income Ratio: Assume WIID statistics exclude nonrespondents, with respondents' mean income at least 75% of the overall mean income but the distribution of this income among the nonrespondents is arbitrary.

3.10 Figures

Fig. 3.1 (3.1)

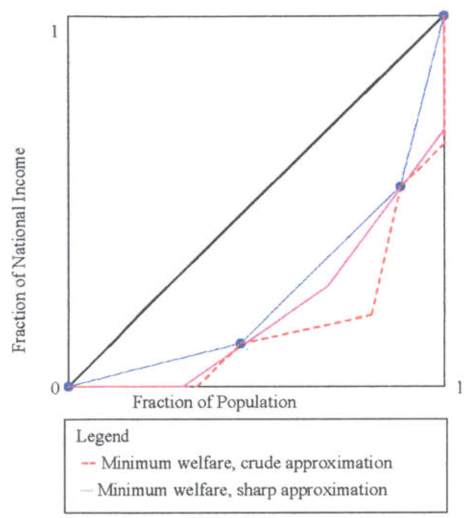


Figure 3.2 (3.2)

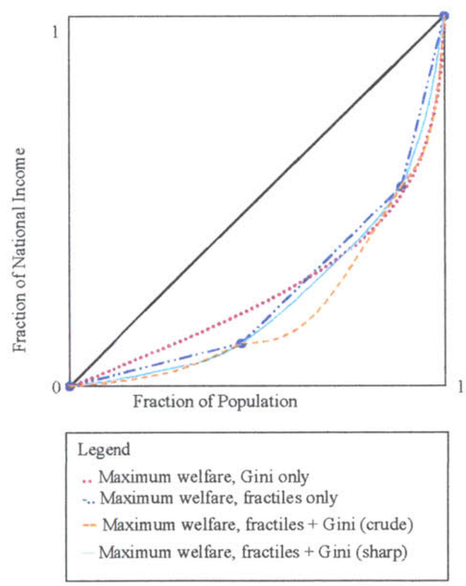


Figure 3.3 (3.3)

Figure 3.4 (3.4)

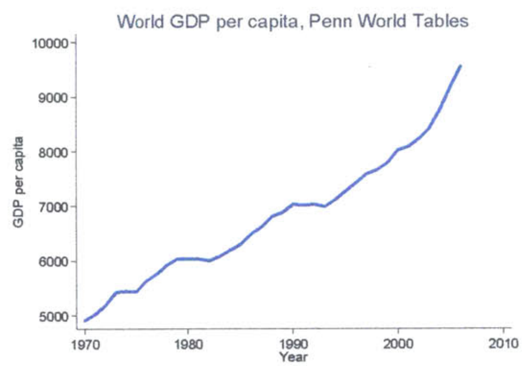


Fig. 3.5 (3.5)

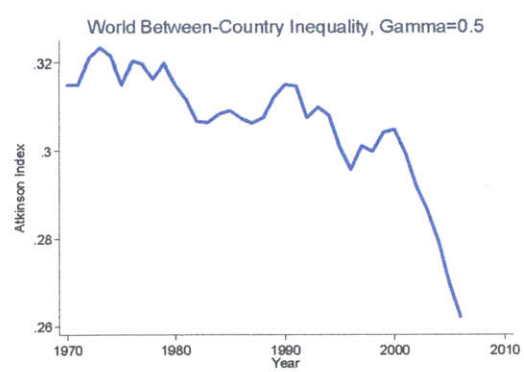
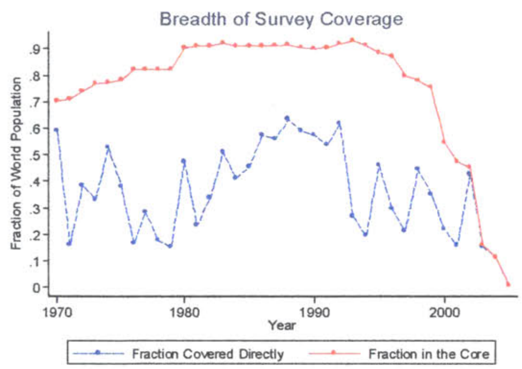


Fig. 3.6 (3.6)



Note: The source for Figures 3.4 and 3.5 is the Penn World Tables 7. The source for Figure **3.6** is the WIID inequality database. In Figure **3.6**, fraction covered directly is fraction of world population in given year in countries with surveys used from WIID. Fraction in core is fraction of world population in given year in countries with at least one earlier survey and at least one later survey.

Figure 3.7 (3.7)

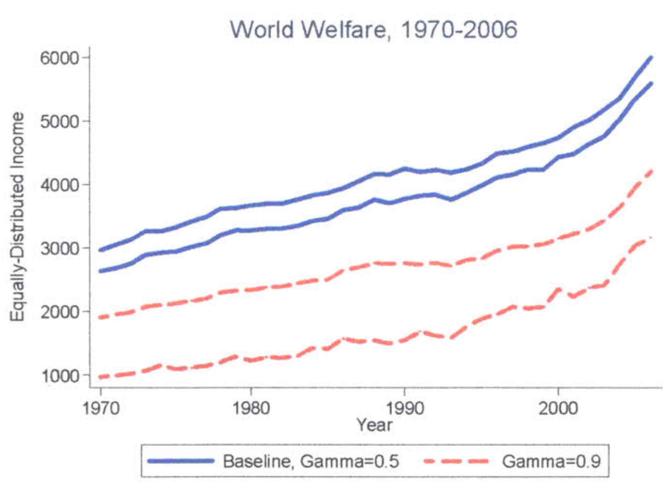
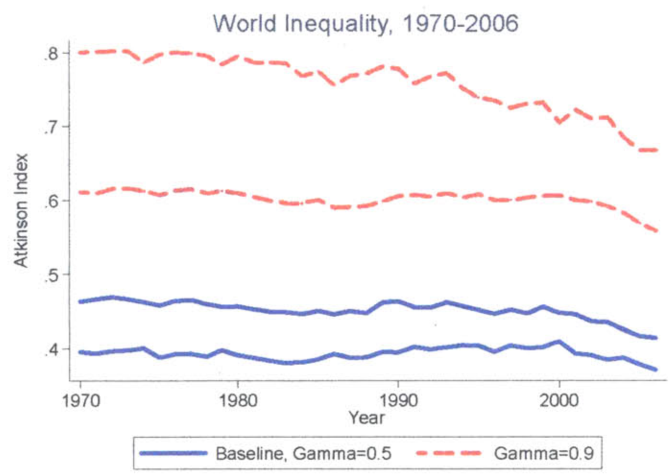
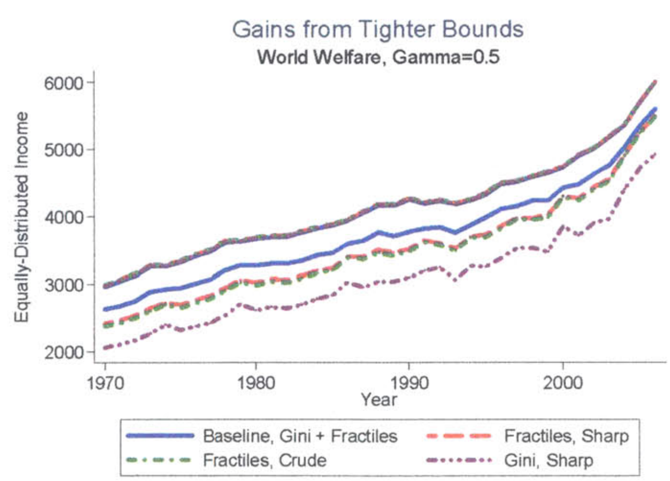


Figure 3.8 (3.8)



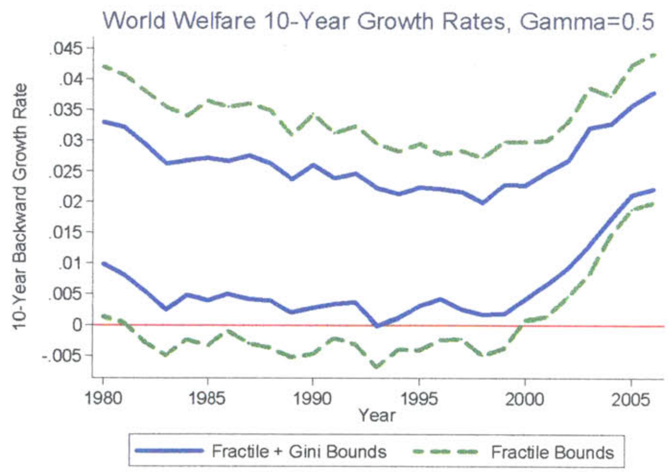
Note: Figure 3.7 (resp., Figure 3.8) shows the bounds on global Atkinson welfare indices (resp., global Atkinson inequality indices) for parameters $\gamma = 0.5$ and $\gamma = 0.9$. Each pair of identically formatted lines represents a pair of bounds. The bounds are sharp given the fractile shares and Gini coefficient for any household survey used, subject to interpolation, extrapolation and imputation for countries and years without inequality data. Mean incomes to construct Figure 3.7 are taken from the Penn World Tables 7.

Fig. 3.9 (3.9)



Note: Figure 3.9 presents sharp global Atkinson welfare bounds based on different statistics from the household surveys in the WIID database. Note that most of the upper bounds coincide.

Fig. 3.10 (3.10)



Note: Figure 3.10 presents sharp bounds on global Atkinson welfare growth based on different statistics from the household surveys in the WIID database.

Fig. 3.11 (3.11)

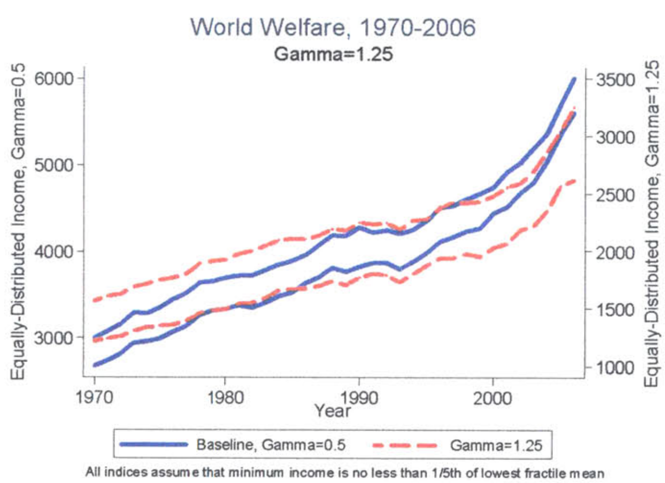
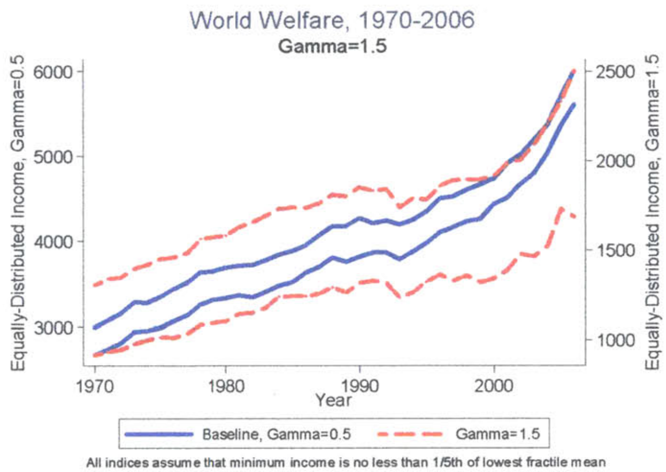


Fig.3.12 (3.12)



Note: Figures 3.11 and 3.12 presents sharp global Atkinson welfare bounds for higher values of the parameter γ. Different series have their own y-axes because of the difference in scales in the indices.

Fig. 3.13 (3.13)

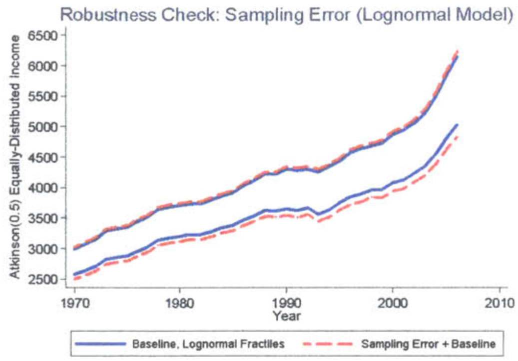


Fig. 3.14 (3.14)

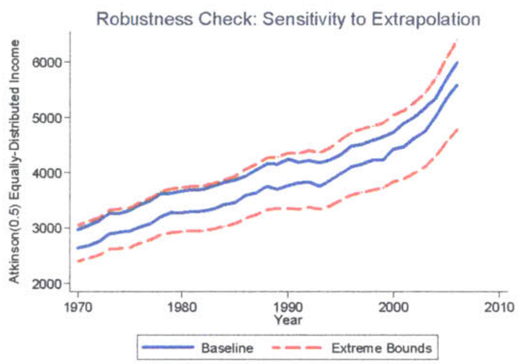


Fig. 3.15 (3.15)

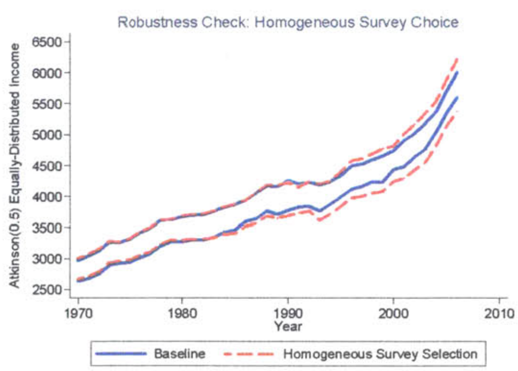


Fig. 3.16 (3.16)

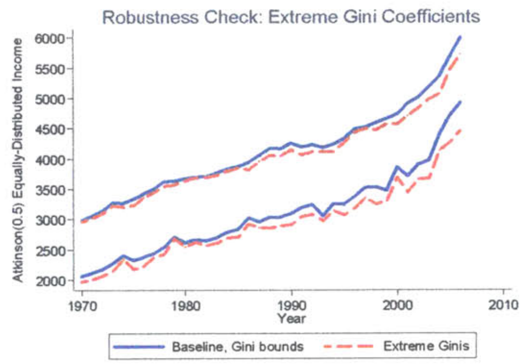


Fig. 3.17 (3.17)

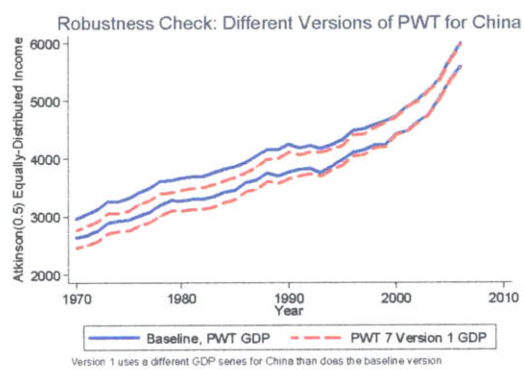
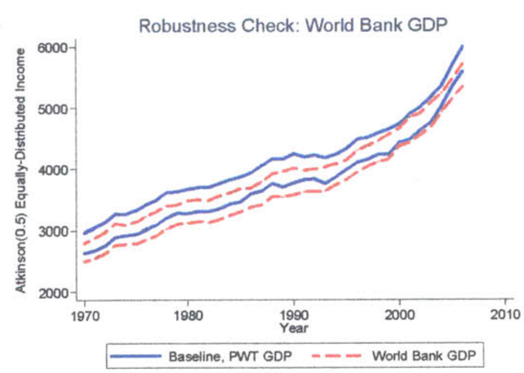


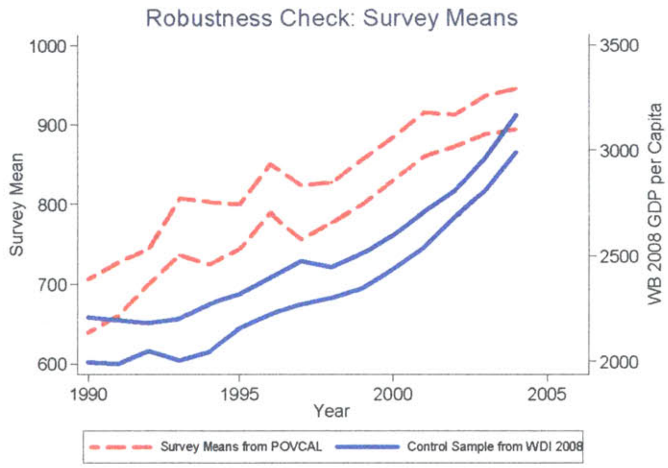
Fig. 3.18 (3.18)



Note: Figure 3.13 checks robustness with respect to sampling error by assuming the underlying inequality distributions in WIID are lognormal and comparing the bounds based on the predicted quintile shares with bounds that incorporate sampling error in the quintile shares. Figure 3.14 checks for robustness to imputing country and regional extreme values of the Atkinson welfare index for countries and years with missing data instead of extrapolation. Figure 3.15 checks for robustness to constraining all surveys selected from WIID for a given country to come from the same source. Figure 3.16 checks for robustness to using the envelope of the bounds for the highest and lowest Gini coefficient provided for every country-year with inequality data;

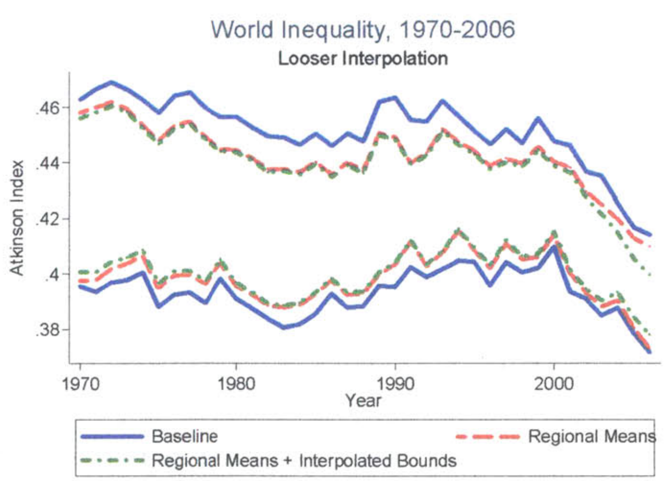
the baseline series is also based only on Gini coefficients. Figure 3.17 checks for robustness to using GDP from the 2008 World Development Indicators. Figure 3.18 checks for robustness to using an alternative GDP series for China reported in the Penn World Tables.

Fig. 3.19 (3.19)



Note: Figure 3.19 checks robustness for the assumption that survey means provide a better measure for mean income than do national accounts GDP. I compute bounds for a sample of 79 countries, which includes the largest and most populous countries in the developing world. The series in red presents bounds computed using survey means from the World Bank's PovCalNet website, while the series in blue presents bounds computed using PWT 7 GDP. The left vertical axis is to be used with the survey means series, while the right vertical axis is to be used with the national accounts series.

Fig. 3.20 (3.20)



Note: Figure 3.20 shows sharp bounds for global Atkinson inequality indices for three different assumptions about interpolation, extrapolation and imputation of the country Atkinson inequality index bounds when inequality data for the given country-years is not available. The three different assumptions are:

Baseline: interpolate the bounds as the outer envelope of the bounds of the adjacent years with survey data. Extrapolate the bounds as the outer envelope of horizontal and linear extrapolation of each bound. Impute bounds for countries with one or no surveys as the outer envelope of the bounds for all countries in the given region at any time in the sample period.

Regional Means: same as Baseline, but impute upper (lower) bound for countries with one or no surveys as the average of the upper (lower) bounds for all countries in the given region in that year.

Regional Means plus Interpolated Bounds: same as Regional Means, but for countries with more than one survey, interpolate each bound linearly and extrapolate the bounds horizontally.

Fig. 3.21 (3.21)

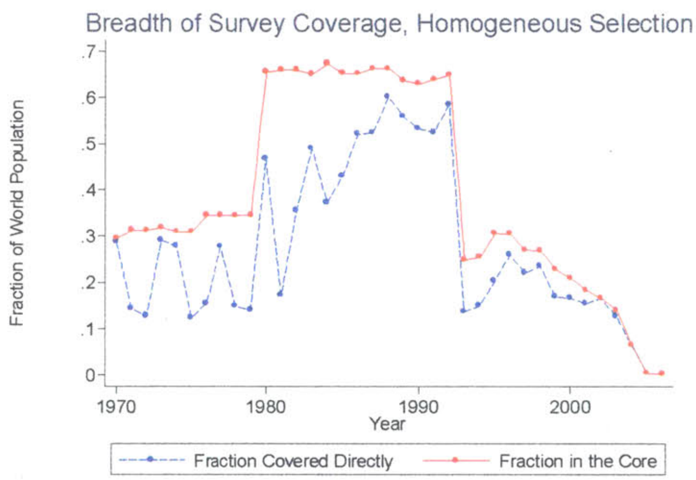
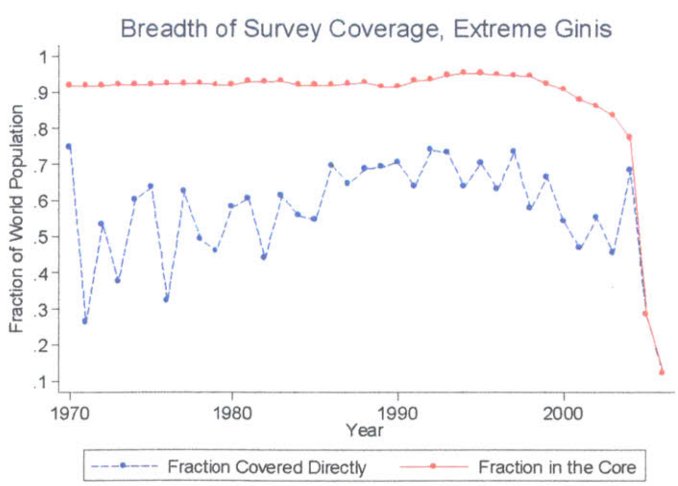


Fig. 3.22 (3.22)



Note: Figure 3.21 shows breadth of coverage when selecting only surveys from the same source from the WIID database; Figure 3.22 shows breadth of coverage when selecting all surveys with national coverage from WIID. Fraction covered directly is fraction of world population in given year in countries with surveys used from WIID according to the selection procedure. Fraction in core is fraction of world population in given year in countries with at least one earlier survey and at least one later survey selected by the procedure.

Fig. 3.23 (3.23)

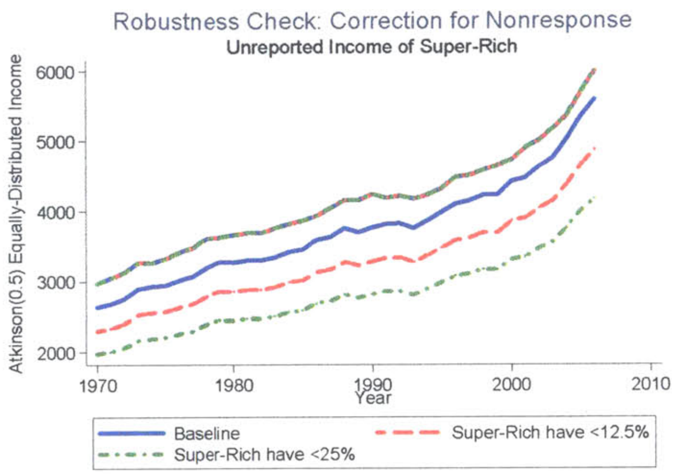
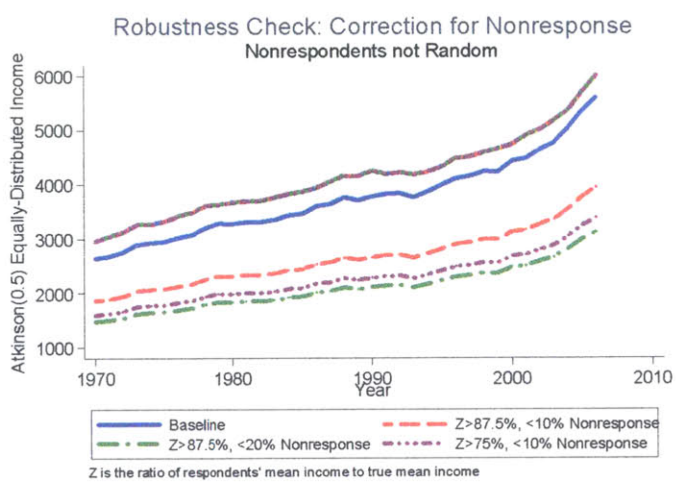


Fig. 3.24 (3.24)



Note: Figure Xa checks robustness for the assumption that the WIID statistics exclude a set of people of measure zero who own <12.5% and <25% of GDP respectively. Figure Xb checks robustness for the assumption that the WIID statistics exclude nonrespondents, and that the ratio between respondents' mean income and true mean income is Z.

Bibliography

[1] Acemoglu, Daron, and Amy Finkelstein. 2008. "Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector." Journal of Political Economy 116, no. 5: 837-880.

[2] Acemoglu, Daron, and Melissa Dell. 2010. "Productivity Differences between and within Countries." American Economic Journal: Macroeconomics 2, no. 1: 169-188.

[3] Acemoglu, Daron, and Simon Johnson. 2005. "Unbundling Institutions." Journal of Political Economy 113, no. 5: 949-995.

[4] Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2000. "The Colonial Origins of Comparative Development: An Empirical Investigation." NBER Working Papers: 7771.

[5] Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." American Economic Review 91, no. 5: 1369-1401.

[6] Acemoglu, Daron, Simon Johnson, James A. Robinson and Pierre Yared. 2008. "Income and Democracy." American Economic Review 98, no.3: pp. 808-42

[7] Acemoglu, Daron. 2005. "Politics and Economics in Weak and Strong States." Journal of Monetary Economics 52:1199-1226.

[8] Ahluwalia, Montek S., Nicholas G. Carter, and Hollis B. Chenery. "Growth and Poverty in Developing Countries." Journal of Development Economics 6, no. 3: 299-341. 1979.

[9] Aitchitson J. and J.A.C. Brown. The Lognormal Distribution. Cambridge, UK: Cambridge University Press. 1957

[10] Algan, Yann, and Pierre Cahuc. 2010. "Inherited Trust and Growth." American Economic Review 100, no. 5: 2060-2092.

[11] Anderson, T. W., and Cheng Hsiao. 1982. "Formulation and Estimation of Dynamic Models Using Panel Data." Journal of Econometrics 18, no. 1: 47-82

[12] Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." Review of Economic Studies 58, no. 2: 277-297.

[13] Arrow, Kenneth J. 1963. "Uncertainty and the welfare economics of medical care." American Economic Review 53:941-73.

[14] Atkinson, Anthony B. "On the Measurement of Inequality." Journal of Economic Theory 2, no. 3: 244-263. 1970.

[15] Atkinson, Anthony B., and Andrea Brandolini. "On Analysing the World Distribution of Income," forthcoming in the World Bank Economic Review. 2010.

[16] Atkinson, Anthony B., and Andrea Brandolini. "Promise and Pitfalls in the Use of 'Secondary' Data-Sets: Income Inequality in OECD Countries As a Case Study." Journal of Economic Literature 39, no. 3: 771-799. 2001.

[17] Atkinson, Anthony B., Thomas Piketty and Emmanuel Saez. "Top Incomes in the Long Run of History." NBER Working Paper #15408. 2009.

[18] Baker, Laurence C., and Ciaran S. Phibbs. 2002. "Managed Care, Technology Adoption, and Health Care: The Adoption of Neonatal Intensive Care." RAND Journal of Economics 33, no. 3: 524-548.

[19] Baker, Laurence C., and Joanne Spetz. 1999. "Managed Care and Medical Technology Growth." In Frontiers in health policy research. Volume 2, 27-52. n.p.: Cambridge and London:,

[20] Baker, Laurence C., and Susan K. Wheeler. 1998. "Managed Care and Technology Diffusion: The Case of MRI." Health Affairs 17, no. 5: 195-207.

[21] Baker, Laurence. C. 1997. "The effect of HMOs on fee-for-service health care expenditures: Evidence from Medicare." Journal of Health Economics, 16(1), 453 481.

[22] Banerjee, Abhijit, and Lakshmi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." American Economic Review 95, no. 4: 1190-1213.

[23] Banerjee, Abhijit, and Thomas Piketty. "Top Indian Incomes, 1922-2000." World Bank Economic Review 19, no. 1: 1-20. 2005.

[24] Barro, Robert J., and Jong-Wha Lee. 2010. "A New Data Set of Educational Attainment in the World, 1950-2010." NBER Working Papers: 15902.

[25] Baugh, Kimberly E., Christopher D. Elvidge , Tilottama Ghosh and Daniel Ziskin. 2009. "Development of a 2009 Stable Lights Product using DMSP-OLS Data." Proceedings of the 30th Asia-Pacific Advanced Network Meeting.

[26] Bester, Alan C., Timothy G. Conley, Christian B. Hansen and Timothy J. Vogelsang. "Fixed-b Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators."

[27] Bhalla, Surjit S. Imagine there's no country: Poverty, inequality, and growth in the era of globalization. Washington, D.C.:Penguin. 2002.

[28] Blendon, R. J. M Brodie, J M Benson, D E Altman, L Levitt, T Hoff and L Hugick. 1998. "Understanding the managed care backlash." Health Affairs, 17, no.4:80-94

[29] Bloche, Gregg M. and David M. Studdert. "A Quiet Revolution: Law as an Agent of Health System Change." Health Affairs 23, no 2: 29-42.

[30] Blundell, Richard, and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." Journal of Econometrics 87, no. 1: 115-143

[31] Bourguignon, Francois, and Christian Morrisson. "Inequality among World Citizens: 1820-1992." American Economic Review 92, no. 4: 727-744. 2002.

[32] Brodie, M. L A Brady and D E Altman. 1998.Media coverage of managed care: is there a negative bias? Health Affairs, 17, no.1:9-25

[33] Buchmueller, Thomas C., and John DiNardo. 2002. "Did Community Rating Induce an Adverse Selection Death Spiral? Evidence from New York, Pennsylvania, and Connecticut." American Economic Review 92, no. 1: 280-294.

[34] Buchmueller, Thomas C., and Su Liu. 2005. "Health Insurance Reform and HMO Penetration in the Small Group Market." Inquiry 42, no. 4: 367-380.

[35] Center for Disease Control. 2012. Compressed Mortality File: http://wonder.cdc.gov/mortSQL.html. Accessed April 21, 2012.

[36] Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT). 2005. Gridded Population of the World, Version 3 (GPWv3). Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw.

[37] Centers for Medicare & Medicaid Services (2011b). National Health Expenditure Data. Retrieved April 21, 2012 from http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/prov-tables.pdf

[38] Chen, Shaohua, and Martin Ravallion. "How Did the World's Poorest Fare in the 1990s?." Review of Income and Wealth 47, no. 3: 283-300. 2001.

[39] Chen, Shaohua, and Martin Ravallion. "The Developing World Is Poorer Than We Thought, but No Less Successful in the Fight against Poverty." Quarterly Journal of Economics 125, no. 4: 1577-1625. 2010.

[40] Chen, Xi, and William D. Nordhaus. 2010. "The Value of Luminosity Data as a Proxy for Economic Statistics." NBER Working Papers: 16317.

[41] Chernew, M. E., DeCicca, P., & Robert, T. 2008. "Managed care and medical expenditures of Medicare beneficiaries." Journal of Health Economics, 27(6), 1451 1461.

[42] Chernew, Michael E. and Joseph P. Newhouse. 2011. "Health Care Spending Growth." in Handbook of Health Economics. Volume 2, 1-43. Handbooks in Economics, vol. 17.

[43] Chiappori, Pierre-Andre, and Monica Paiella. "Relative Risk Aversion is Constant: Evidence from Panel Data," forthcoming Journal of the European Economic Association. 2006.

[44] Chotikapanich, Duangkamon, William E. Griffiths, and D. S. Prasada Rao. "Estimating and Combining National Income Distributions Using Limited Data." Journal of Business and Economic Statistics 25, no. 1: 97-109. 2007.

[45] Clemens, Jeffrey and Joshua Gottlieb. 2012. Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health? mimeo, Havard.

[46] Conley, Timothy. G. 1999. "GMM Estimation with Cross Sectional Dependence." Journal of Econometrics 92, no. 1: 1-45.

[47] Cowell, Frank A. "Grouping Bounds for Inequality Measures under Alternative Informational Assumptions." Journal of Econometrics 48, no. 1-2: 1-14. 1991.

[48] Cowell, Frank A. "Inequality Decomposition: Three Bad Measures." Bulletin of Economic Research 40, no. 4: 309-312. 1988.

[49] Cowell, Frank A. "Measuring Inequality" in Atkinson T. and Bourguignon F. eds. Handbook of Income Distribution, Elsevier. 2000

[50] Cowell, Frank A. Measuring Inequality, Oxford: Philip Allan. 1977

[51] Cressie, Noel G. 1993. Statistics for Spatial Data. New York: J. Wiley.

[52] Currie, Janet, and Jonathan Gruber. 1996. "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women." Journal of Political Economy 104, no. 6: 1263-1296.

[53] Cutler, David M. 2004. Your Money or Your Life: Strong Medicine for America's Health Care System. New York: Oxford University Press.

[54] Cutler, David M. 2010. "Where Are the Health Care Entrepreneurs? The Failure of Organizational Innovation in Health Care." In Innovation Policy and the Economy. Volume 11, 1-28. Chicago and London: University of Chicago Press.

[55] Cutler, David M., and Louise Sheiner. 1997. "Managed Care and the Growth of Medical Expenditures." NBER Working Paper #6140.

[56] Cutler, David M., Mark McClellan, and Joseph P. Newhouse. 2000. "How Does Managed Care Do It?." RAND Journal of Economics 31, no. 3: 526-548.

[57] Dafny, Leemore S. 2005. "How Do Hospitals Respond to Price Changes?." American Economic Review 95, no. 5: 1525-1547.

[58] de Soto, Hernando. 2000. The Mystery of Capital. New York: Basic Books.

[59] Deaton, Angus. "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)." Review of Economics and Statistics 87, no. 1: 1-19. 2005.

[60] Deaton, Angus. "Price Indexes, Inequality, and the Measurement of World Poverty." American Economic Review 100, no. 1: 5-34. 2010.

[61] Declercq, Eugene and Diana Simmes. 1997. "The Politics of "Drive-through Deliveries": Putting Early Postpartum Discharge on the Legislative Agenda." The Milbank Quarterly, Vol. 75, No. 2, pp. 175-202.

[62] Deininger, Klaus, and Lyn Squire. "A New Data Set Measuring Income Inequality," World Bank Economic Review, X, 565–591. 1996

[63] Dell, Melissa. 2010. "The Persistent Effects of Peru's Mining Mita." Econometrica 78, no. 6: 1863-1903.

[64] Dikhanov, Yuri, and Michael Ward. "Evolution of the Global Distribution of Income, 1970–99," mimeo. 2001.

[65] Doll, Christopher N.H. 2008. "CIESIN Thematic Guide to Night-time Light Remote Sensing and its Applications." Manuscript.

[66] Doll, Christopher N.H., Jan Peter Muller and Jeremy G. Morley. 2006. "Mapping Regional Economic Activity from Night-Time Light Satellite Imagery." Ecological Economics 57: 75-92

[67] Duggan, Mark and Tamara Hayford. 2011. "Has the Shift to Managed Care Reduced Medicaid Expenditures? Evidence from State and Local-Level Mandates." NBER Working Paper #17236.

[68] Elvidge, Christopher D. and Kimberly E. Baugh, Eric A. Kihn, Herbert W. Kroehl, Ethan R. Davis. 1997. "Mapping City Lights With Nighttime Data from the DMSP Operational Linescan System." Photogrammetric Engineering & Remote Sensing 63, no. 6: 727-734

[69] Elvidge, Christopher D., Kimberly E. Baugh, John B. Dietz, Theodore Bland, Paul C. Sutton and Herbert W. Kroehl. 1999. "Radiance Calibration of DMSP-OLS Low-Light Imaging Data of Human Settlements." Remote Sensing of Environment, 68(1):77-88.

[70] Elvidge, Christopher D., Kimberly E. Baugh, Sharolyn. J. Anderson, Paul. C. Sutton, and Tilottama Ghosh. 2012. "The Night Light Development Index (NLDI): A Spatially Explicit Measure of Human Development from Satellite Data." Social Geography 7, 23–35.

[71] Elvidge, Christopher. 2003. "Nighttime Lights Change Detection." PowerPoint Presentation.

[72] Feldstein, Paul J. 1988. Health care economics. 3rd ed. New York: Wiley.

[73] Finkelstein, Amy. 2007. "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare." Quarterly Journal of Economics 122, no. 1: 1-37.

[74] Galbraith, James K., and Hyunsub Kum. "Estimating the Inequality of Household Incomes: A Statistical Approach to the Creation of a Dense and Consistent Global Data Set." Review of Income and Wealth 51, no. 1: 115-143. 2005.

[75] Gastwirth, Joseph L. "The Estimation of the Lorenz Curve and Gini Index." Review of Economics and Statistics 54, no. 3: 306-316. 1972.

[76] Gawande, Atul. 2009. "The Cost Conundrum." The New Yorker.

[77] Gennaioli, Nicola, and Ilia Rainer. 2007. "The Modern Impact of Precolonial Centralization in Africa." Journal of Economic Growth 12, no. 3: 185-234.

[78] Ghosh, Tilottama, Rebecca L. Powell, Christopher D. Elvidge, Kimberly E. Baugh, Paul C. Sutton and Sharolyn Anderson. 2010. "Shedding Light on the Global Distribution of Economic Activity." The Open Geography Journal 3, 148-161.

[79] Gibrat, R. Les inégalités économiques, Paris: Sirey. 1931.

[80] Ginsburg, P. B. & Pickreign, J. D. 1997. "Tracking health care costs: An update." Health Affairs, 16(4), 151-155.

[81] Ginsburg, P. B. & Pickreign, J.D. 1996. "Tracking health care costs." Health Affairs, 15(3), 140 149.

[82] Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?." Journal of Economic Growth 9, no. 3: 271-303.

[83] Glied, Sherry. 2000. "Managed Care." in Handbook of Health Economics Volume 1A, 707-753. Handbooks in Economics, vol. 17.

[84] Glied, Sherry. 2003. "Health Care Costs: On the Rise Again." Journal of Economic Perspectives 17, no. 2: 125-148.

[85] Gray, Virginia, David Lowery and Erik K. Godwin. 2007. "The Political Management of Managed Care: Explaining Variations in State Health Maintenance Organization Regulations." Journal of Health Politics, Policy and Law 32 no. 3.

[86] Gruber, Jonathan, and Kosali Simon. 2008. "Crowd-Out 10 Years Later: Have Recent Public Insurance Expansions Crowded Out Private Health Insurance?." Journal of Health Economics 27, no. 2: 201-217.

[87] Gruber, Jonathan. 1994. "The Incidence of Mandated Maternity Benefits." American Economic Review 84, no. 3: 622-641.

[88] Gruber, Jonathan. 1998. "Health Insurance and the Labor Market." NBER Working Paper #6762.

[89] Gruber, Jonathan. 2010. "The Tax Exclusion for Employer-Sponsored Health Insurance." NBER Working Paper #15766.

[90] Hahn, Jinyong, Jerry Hausman, and Guido Kuersteiner. 2007. "Long Difference Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects." Journal of Econometrics 140, no. 2: 574-617.

[91] Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." Econometrica 69, no. 1: 201-209.

[92] Hall, Robert E., and Charles I. Jones. 2004. "The Value of Life and the Rise in Health Spending." NBER Working Paper #10737.

[93] Hausman, Jerry A. and Maxim L. Pinkovskiy. 2013. "A Nonlinear Least Squares Approach to Estimating Fixed Effects Panel Data Models with Lagged Dependent Variables." In progress, MIT.

[94] Henderson, Vernon, Adam Storeygard, and David N. Weil. 2011. "A Bright Idea for Measuring Economic Growth." American Economic Review 101, no. 3: 194-199

[95] Heston, Alan, Robert Summers and Bettina Aten. 2011. Penn World Table Version 7.0, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, May 2011.

[96] Heston, Allan, Robert Summers, and Bettina Aten. Penn World Table Version 6.2, Center for International Comparisons at the University of Pennsylvania (CICUP). 2006

[97] Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis. 2005. "Very high resolution interpolated climate surfaces for global land areas." International Journal of Climatology 25: 1965-1978.

[98] Huillery, Elise. 2009. "History Matters: The Long-Term Impact of Colonial Public Investments in French West Africa." American Economic Journal: Applied Economics 1, no. 2: 176-215.

[99] Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." Journal of Econometrics 142, no. 2: 615-635.

[100] Inglehart, Ronald and Christian Welzel. 2008. "The WVS Cultural Map of the World." http://www.worldvaluessurvey.org/wvs/articles/folder_published/article_base_54. Accessed March 9, 2013.

[101] International Monetary Fund. 2005. Direction of Trade Statistics, Database and Browser CD-ROM. Washington, D.C.

[102] Jacobson, Peter D. and Shannon Brownlee. 2005. "The Health Insurance Industry and the Media: Why the Insurers Aren't Always Wrong." 5 Houston Journal of Health Law & Policy 235.

[103] Johnson, Simon, William Larson, Chris Papageorgiou, and Arvind Subramanian. 2009. "Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates." NBER Working Papers: 15455

[104] Kaiser Family Foundation and Health Research & Education Trust. 2011. "Employer Health Benefits: 2011 Annual Survey." Full Report.

[105] Kalecki, M. "On the Gibrat Distribution", Econometrica, 13, 161-170. 1945.

[106] Kamien, Morton I. and Nancy L. Schwartz. Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management. New York: Elsevier. 1991.

[107] La Porta, Rafael Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny. 1998. "Law and Finance." Journal of Political Economy 106, no. 6: 1113-1155.

[108] La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." Journal of Economic Literature 46, no. 2: 285-332.

[109] Larreguy, Horacio. 2011. "The Lasting Impact of Colonial Policies in Nigeria: Evidence from a Natural Experiment." Manuscript.

[110] Laudicina, Susan et al. 2011. "State Legislative Health Care and Insurance Issues." Blue Cross Blue Shield Association of America.

[111] Lee, David S., and David Card. 2008. "Regression Discontinuity Inference with Specification Error." Journal of Econometrics 142, no. 2: 655-674.

[112] Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." Journal of Economic Literature 48, no. 2: 281-355.

[113] Lopez, J. Humberto and Servén, Luis. "A Normal Relationship? Poverty, Growth and Inequality," World Bank Policy Research Working Paper 3814. 2006.

[114] Maddison, Angus. Historical Statistics of the World Economy: 1-2003 A.D. http://www.ggdc.net/maddison/ Date Accessed: June 1 2007.

[115] Manski, Charles. Identification Problems in the Social Sciences. Cambridge, MA: Harvard University Press. 1995

[116] Matheron, G. 1963. "Principles of Geostatistics." Economic Geology 58:1246-1266

[117] McDonald, James B., and Michael R. Ransom. "An Analysis of the Bounds for the Gini Coefficient." Journal of Econometrics 17, no. 2: 177-188. 1981.

[118] McDonald, James B., and Yexiao J. Xu. "A Generalization of the Beta Distribution with Applications." Journal of Econometrics 66, no. 1-2: 133-152. 1995.

[119] Mehran, Farhad. "Bounds on the Gini Index Based on Observed Points of the Lorenz Curve." Journal of the American Statistical Association 70, no. 349: 64-66. 1975.

[120] Michalopoulos, Stelios and Elias Papaioannou. 2011a. "Divide and Rule or the Rule of the Divided? Evidence from Africa." NBER WP 17184

[121] Michalopoulos, Stelios and Elias Papaioannou. 2011b. "The Long-Run Effects of the Scramble for Africa." Manuscript.

[122] Milanovic, Branko. "Global Inequality Recalculated and Updated: The Effect of New PPP Estimates on Global Inequality and 2005 Estimates." Journal of Economic Inequality 10, no. 1: 1-18. 2012.

[123] Milanovic, Branko. "True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone." Economic Journal 112, no. 476: 51-92. 2002.

[124] Milanovic, Branko. Worlds Apart: Measuring International and Global Inequality. Princeton and Oxford: Princeton University Press. 2005.

[125] Miller, Robert H. and Harold S. Luft. 1997. "Does managed care lead to better or worse quality of care?" Health Affairs, 16, no.5:7-25

[126] Miller, Robert H. and Harold S. Luft. 2002. "HMO Plan Performance Update: An Analysis of The Literature, 1997-2001." Health Affairs, 21, no.4 :63-86

[127] Murray, David. "Extreme Values for the Gini Coefficients Calculated from Grouped Data." Economics Letters 1: 389-393. 1978.

[128] National Council of State Legislatures. 2011. "Managed Care State Laws and Regulations, Including Consumer and Provider Protections." http://www.ncsl.org/issues-research/health/managed-care-state-laws.aspx

[129] Newhouse, Joseph P. 1992. "Medical Care Costs: How Much Welfare Loss?" Journal of Economic Perspectives 6, no. 3: 3-21.

[130] Newman, Charles M. 1984. "Asymptotic Independence and Limit Theorems for Positively and Negatively Dependent Random Variables." Lecture Notes-Monograph Series, Vol. 5, Inequalities in Statistics and Probability: 127-140

[131] Nunn, Nathan, and Diego Puga. 2009. "Ruggedness: The Blessing of Bad Geography in Africa." NBER Working Papers: 14918.

[132] Nunn, Nathan. 2008. "The Long-Term Effects of Africa's Slave Trades." Quarterly Journal of Economics 123, no. 1: 139-176.

[133] Pakes, Ariel, and Zvi Griliches. 1980. "Patents and R and D at the Firm Level: A First Look." NBER Working Papers: 0561.

[134] Pareto, Vilfredo. "Cours d'Economie Politique", F. Rouge, Lausanne. 1897.

[135] Peterson, Mark A. "Introduction, Politics, Misperception or Apropos?" Journal of Health Policy, Politics and Law 24, no. 5: 873-886.

[136] Pinkovskiy, Maxim and Xavier Sala-i-Martin. "Parametric Estimations of the World Distribution of Income." NBER Working Paper #15433. 2009.

[137] Porter, Jack. 2003. "Estimation in the Regression Discontinuity Model." Manuscript.

[138] Ravallion, Martin, Johan A. Mistiaen, and Anton Korinek. "Survey nonresponse and the distribution of income." World Bank Policy Research Working Paper 3543. 2005.

[139] Reddy, Sanjay G., and Camelia Minoiu. "Has World Poverty Really Fallen?" Review of Income and Wealth 53, no. 3: 484-502. 2007.

[140] Robinson, Peter M. 2011. "Asymptotic Theory for Nonparametric Regression with Spatial Data." Journal of Econometrics 165, no.1: 5-19

[141] Sala-i-Martin, Xavier. "The Disturbing 'Rise' of Global Income Inequality." NBER Working Paper #8904. 2002a.

[142] Sala-i-Martin, Xavier. "The World Distribution of Income (estimated from Individual Country Distributions)." NBER Working Paper #8933. 2002b.

[143] Sala-i-Martin, Xavier. "The World Distribution of Income: Falling Poverty and . . . Convergence, Period." Quarterly Journal of Economics 121, no. 2: 351-397. 2006.

[144] Salem, A. and T. Mount. "A Convenient Descriptive Model of Income Distribution: The Gamma Density", Econometrica 42, 1115-1128. 1974.

[145] Schabenberger, Oliver and Carol Gotway. 2005. Statistical Methods for Spatial Data Analysis. Boca Raton: Chapman and Hall/CRC

[146] Schultz, T. Paul. "Inequality in the Distribution of Personal Income in the World: How It Is Changing and Why." Journal of Population Economics 11, no. 3: 307-344. 1998.

[147] Sen, Amartya K. "Real National Income." Review of Economic Studies 43, no. 1: 19-39. 1976.

[148] Simon, Kosali Ilayperuma. 2005. "Adverse Selection in Health Insurance Markets? Evidence from State Small-Group Health Insurance Reforms." Journal of Public Economics 89, no. 9-10: 1865-1877.

[149] Singh, S. and G. Maddala. "A Function for Size Distribution of Incomes", Econometrica 44, 963-970. 1976.

[150] Skinner JS, Staiger DO, Fisher ES. 2006. "Is Technological Change in Medicine Always Worth It? The Case of Acute Myocardial Infarction." Health Affairs web exclusive.

[151] Socio-Economic Database for Latin America and the Caribbean (CEDLAS and The World Bank). Date Accessed: September 1, 2009.

[152] Spolaore, Enrico and Romain Wacziarg. 2009. "The Diffusion of Development," Quarterly Journal of Economics, 124 no.2: pp 469-592

[153] Statistical Abstract of the United States: http://www.census.gov/compendia/statab/. Accessed on April 21, 2012.

[154] Stein, Michael L. 1987. "A Modification of Minimum Norm Quadratic Estimation of a Generalized Covariance Function for Use with Large Data Sets." Journal of the American Statistical Association 82, no. 399: 765-772

[155] Stein, Michael L. 1999. Interpolation of spatial data : some theory for kriging. New York: Springer.

[156] Sutton, Paul C., Christopher D. Elvidge and Tilottama Ghosh. 2007. "Estimation of Gross Domestic Product at Sub-National Scales using Nighttime Satellite Imagery." International Journal of Ecological Economics & Statistics 8:5-21

[157] Thomasson, Melissa A. 2003. "The Importance of Group Coverage: How Tax Policy Shaped U.S. Health Insurance." American Economic Review 93, no. 4: 1373-1384.

[158] U.S. National Imagery and Mapping Agency (NIMA). 2003. VMAP0DATA National Stock Number (NSN): 7644014361830, 7644014361831, 7644014361833 . Fairfax, VA.

[159] USGS. 2000. Shuttle Radar Topography Mission. Global Land Cover Facility, University of Maryland, College Park, Maryland.

[160] Ware Jr., J.E., R.H. Brook, W.H. Rogers, E.B. Keeler, A.R. Davies, C.D. Sherbourne, G.A. Goldberg, P. Camp and J.P. Newhouse. 1987. "Health outcomes for adults in prepaid and fee-for-service systems of care: results from the health insurance experiment", RAND R-3459-HHS (RAND Corporation, Santa Monica, CA).

[161] Weisbrod, Burton A. 1991. "The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment." Journal of Economic Literature 29, no. 2: 523-552.

[162] World Bank. 2011. World Development Indicators. Washington, DC.

[163] World Bank. 2011. World Governance Indicators. Washington, DC.

[164] World Bank. World Development Indicators 2009. Accessed July 15, 2007 and June 1, 2008. http://go.worldbank.org/U0FSM7AQ40

[165] Zhu, Feng. "A Nonparametric Analysis of the Shape Dynamics of the U.S. Personal Income Distribution: 1962-2000." BIS Working Paper #184. 2005.