

NBER WORKING PAPER SERIES

NATURAL EXPERIMENTS IN MACROECONOMICS

Nicola Fuchs-Schuendeln
Tarek Alexander Hassan

Working Paper 21228
<http://www.nber.org/papers/w21228>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2015

Chapter prepared for the Handbook of Macroeconomics. The chapter has benefitted from helpful discussions and comments from Daron Acemoglu, Jonathan Dingel, Yuriy Gorodnichenko, Chang-Tai Hsieh, Nathan Nunn, Elias Papaioannou, Matthias Schündeln, Rob Vishny, Mirko Wiederholt, and seminar participants of the Handbook of Macroeconomics Conference at Stanford and at the Deutsche Bundesbank. Leonhard Czerny, Denis Gorea, and Philip Xu provided excellent research assistance. Fuchs-Schündeln gratefully acknowledges financial support from the Cluster of Excellence “Formation of Normative Orders” and the European Research Council under Starting Grant No. 262116. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Nicola Fuchs-Schuendeln and Tarek Alexander Hassan. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Natural Experiments in Macroeconomics
Nicola Fuchs-Schuendeln and Tarek Alexander Hassan
NBER Working Paper No. 21228
June 2015, Revised January 2016
JEL No. C1,C9,E21,E62,H31,O11,O14,O43,O50

ABSTRACT

A growing literature relies on natural experiments to establish causal effects in macroeconomics. In diverse applications, natural experiments have been used to verify underlying assumptions of conventional models, quantify specific model parameters, and identify mechanisms that have major effects on macroeconomic quantities but are absent from conventional models. We discuss and compare the use of natural experiments across these different applications and summarize what they have taught us about such diverse subjects as the validity of the Permanent Income Hypothesis, the size of the fiscal multiplier, and about the effects of institutions, social structure, and culture on economic growth. We also outline challenges for future work in each of these fields, give guidance for identifying useful natural experiments, and discuss the strengths and weaknesses of the approach.

Nicola Fuchs-Schuendeln
Goethe University Frankfurt
House of Finance
60323 Frankfurt
Germany
fuchs@wiwi.uni-frankfurt.de

Tarek Alexander Hassan
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
tarek.hassan@chicagobooth.edu

Contents

1	Introduction	4
2	Verification: The Permanent Income Hypothesis	8
2.1	Reaction of Consumption to Unexpected Income Shocks	9
2.1.1	Unexpected Temporary Income Shocks	10
2.1.2	An Unexpected Permanent Income Shock: the Natural Experiment of German Reunification	11
2.2	Reaction of Consumption to Expected Income Changes	14
2.2.1	Random Treatment: Determining an Appropriate Control Group	16
2.2.2	The Presence of Liquidity Constraints	21
2.2.3	Overview of Natural Experiment Studies of the Permanent Income Hypothesis	24
2.2.4	Violation of Rational Expectations or Need for Model Extension?	25
3	Quantification: The Fiscal Multiplier	30
3.1	Permanent Income Hypothesis Studies and the Fiscal Multiplier	31
3.2	Military News Shocks as Natural Experiments	33
3.3	Local Fiscal Multipliers	35
4	Identification: Causal Factors in Economic Growth	40
4.1	The Fundamental Causes of Growth	40
4.2	Institutions and Political Economy	44
4.2.1	The Effect of Institutions on Growth	45
4.2.2	The Effect of Institutions on Business Cycles and Conflict	49
4.2.3	Persistent Effects of Historical Institutions	50
4.2.4	Determinants and Dynamics of Institutions	53
4.3	Social Structure	57
4.3.1	The Effect of Social Ties on Growth	58
4.3.2	The Effect of Social Ties on Trade and FDI	61
4.3.3	The Effect of Internal Social Structure on Institutions and Growth	66
4.4	Trust and Civic Capital	68
4.4.1	The Effect of Trust on Growth	69
4.4.2	Effect of Trust on Financial Development and Other Aggregates	71
4.4.3	Determinants and Dynamics of Trust	72

4.5 Multiple Equilibria and Path Dependence	75
5 Critical Assessment and Outlook	79

1 Introduction

Establishing causality is a major challenge in economics, especially in macroeconomics, where the direction of various important causal relationships is widely discussed, as illustrated, for example, by large-scale debates about the causal effects of monetary and fiscal policies. Most empirical applications of macroeconomic models focus on matching conditional correlations and improving the fit of models to a set of data moments. Despite substantial advances in this area in recent years, these conditional correlations often cannot identify causal chains. For example, New Keynesian models and real business cycle models can match similar sets of conditional correlations but have very different predictions about the causal effects of fiscal or monetary policies. This lack of identification of clear causal channels is especially troubling regarding providing policy advice.

In applied microeconomic fields, causality is often established by designing laboratory or field experiments. In these types of experiments, the researcher consciously influences the economic environment in a way that allows the establishment of causality. The most prevalent and clearest method in this spirit is to randomly allocate agents into a treatment group and a control group, and then analyze the effect of the treatment by directly comparing the relevant outcome variables between both groups, or the change in the outcome variables of both groups coinciding with the introduction of the treatment in a difference-in-differences approach. Field experiments randomize treatment in a real-world economic environment, whereas laboratory experiments do so in a controlled environment. Both methods are mostly unavailable to macroeconomists for fairly obvious reasons. Because macroeconomics deals with phenomena that affect the economy at large (e.g., economic growth, unemployment, monetary policy, fiscal policy), any field interventions would be very expensive and would have far-reaching consequences because they cannot easily be targeted at a specific small group, making it unlikely that anyone would agree to carry them out. Bringing key features of the economic environment into the laboratory is also complicated in macroeconomics, where the interplay of different agents and markets often plays a key role (see [Duffy \(2008\)](#) for a survey of laboratory experiments in macroeconomics).

Natural experiments are an alternative to field and laboratory experiments. For the purposes of our discussion, *we define natural experiments as historical episodes that provide observable, quasi-random variation in treatment subject to a plausible identifying assumption*. The “natural” in natural experiments indicates a researcher did not consciously design the episode to be analyzed, but can nevertheless use it to learn about

causal relationships. The episode under consideration can be a policy intervention carried out by policy makers (e.g., changes in the tax law), historical episodes that go beyond simple policy measures (e.g., the fall of Communism), or a so-called “natural natural” experiment that arises from natural circumstances (rainfall, earth quakes, etc.). Maybe the most widely exploited natural experiment in the macroeconomics literature is the German separation in 1949 and subsequent reunification in 1989. This episode split a homogeneous population into two parts that lived under vastly different economic and political systems with minimal contact between them, only to be reunited 40 years later. Importantly, one can argue this split was exogenous to preferences, economic conditions, and other factors that would directly predict different economic outcomes after reunification. Thus, the assignment of an individual to East or West Germany at the date of separation can be considered random, as in a field experiment. At the same time, vast micro and macro data are available to analyze the episode. [Fuchs-Schündeln and Schündeln \(2005\)](#) first used this experiment to study the self-selection into occupations according to risk aversion and its effect on precautionary savings. Later applications have studied diverse subjects ranging from endogenous preferences for economic policies ([Alesina and Fuchs-Schündeln, 2007](#)) and the importance of market access ([Redding and Sturm, 2008](#)) to the economic impact of social ties ([Burchardi and Hassan, 2013](#)).

Whereas the main task of a researcher carrying out a laboratory or field experiment lies in designing it in a way that allows causal inference, the main task of a researcher analyzing a natural experiment lies in arguing that in fact the historical episode under consideration resembles an experiment, and in dealing with weaknesses of the ex-post experimental setup that one would have avoided a priori in a designed experiment. To show the episode under consideration resembles an experiment, identifying valid treatment and control groups, that is, arguing the treatment is in fact randomly assigned, is crucial. Establishing such quasi-random treatment requires showing that two groups are comparable along all dimensions relevant for the outcome variable except the one involving the treatment. The methods used to do this are often adapted from the micro-econometric literature on field and laboratory experiments.

The goal of this chapter is to acquaint the reader with the use of natural experiments in macroeconomics, summarize what we have learned from them so far, and distill what makes a successful application of a natural experiment to answer a macroeconomic question. We provide in the conclusion of this chapter a summary of common features that distinguish successful papers that rely on the use of natural experiments. Although every natural experiment is different and thus leads to different challenges, these features can

serve as guidelines for future papers. Moreover, we discuss the embedding of natural experiments into structural models as a promising general avenue for future research, and point out limitations in the use of natural experiments.

Rather than attempt to cover all papers in macroeconomics that feature natural experiments (which would be a formidable task), we instead select three specific lines of enquiry that use natural experiments for three different purposes: to verify underlying model premises (“verification”), to quantify specific policy parameters (“quantification”), and to identify causal mechanisms that operate outside conventional models (“identification”).

The first line is the literature on the Permanent Income Hypothesis. In contrast to the simple Keynesian consumption theory, the Permanent Income Hypothesis assumes agents are rational and forward-looking when making their consumption decisions. Therefore, in addition to current income and current assets, the expected value of future income plays a role in the optimal consumption choice today. This forward-looking behavior can be subjected to a simple test using a preannounced income change: the household should adjust consumption as soon as information about the future income change arrives. By contrast, consumption growth should be unaffected at the time of the implementation of the income change, given that the household knew about it in advance. In this literature, natural experiments serve to identify such preannounced income changes. A finding that households adjust their consumption at the time of implementation of the preannounced income change casts doubts on the fundamental assumption of most micro-founded macroeconomic models that agents are forward-looking in their decision making.

The second line is the literature striving to quantify the fiscal multiplier. The fiscal multiplier is one of the most important policy parameters in the macroeconomics literature. Can the government stimulate the economy via government spending or tax policies? If yes, how large is the effect of a given fiscal policy on GDP per capita? The main challenge in the estimation of the fiscal multiplier lies in identifying changes in fiscal policies that are not motivated by business-cycle considerations. In this context, researchers specifically use natural experiments to identify such exogenous changes in government spending.

These first two lines of literature rely not only on natural experiments, but also on other approaches, for example, instrumental variables approaches in which the instruments are not historical episodes, or vector autoregression (VAR) models with exclusion restrictions. By contrast, the third line of the literature relies almost exclusively on natural experiments to identify the fundamental causes of growth. The goal of this literature

is to identify mechanisms that are absent from standard macroeconomic models. What can explain the vastly different GDP per-capita levels across poor and rich countries? Standard growth models point to human or physical capital accumulation or R&D investment as explanations, but these factors are proximate rather than fundamental causes of growth: why have some countries invested much more than others? The literature on the fundamental causes of growth identifies institutions, social structure, and culture as such fundamental causes. All three of these concepts are largely absent from conventional models of economic growth. Moreover, multiple equilibria can lead to different growth paths despite common initial conditions. Empirically analyzing the fundamental causes of growth is intimately linked to using natural experiments: the “historic episodes” are truly historic here in the sense that they typically come from the distant past and are used to establish causal links by providing quasi-random variation in institutions, social structure, or culture across countries, regions, or time.

Within each of the three lines of literature, we again do not attempt to survey the entire literature on the topic but instead focus on showing how the authors use natural experiments to address research questions arising within each of the three specific contexts, by verifying, quantifying, or identifying causal mechanisms. A common theme across almost all of these applications is that the econometric methods used are fairly simple applications of standard methods, such as OLS, instrumental variables, regression discontinuity, or fixed-effects estimators. Instead, the complexity of many of these papers lies in identifying the episode that generates quasi-random variation, and appropriately dealing with any flaws in nature’s experimental design. In this sense, the most crucial ingredient of many papers using natural experiments is the appropriate statement and defense of an identifying assumption, which is the focus of our discussion.

This chapter has two target audiences: the first is researchers with a solid background in applied econometrics who are considering studying a natural experiment in any area of macroeconomics. We hope the juxtaposition of natural experiments used in different areas will generate ideas for intellectual arbitrage for this group. In each of the areas that we cover, we also attempt to point out the research frontier in terms of method and substance, and often explicitly point out important avenues for future research. The second target audience is researchers in mainstream macroeconomics. With this group in mind, we attempt to summarize what natural experiments have taught us about the permanent income hypothesis, the fiscal multiplier, and the fundamental causes of macroeconomic growth, in the hope that this summary will help direct future theoretical research.

A set of slides that develops the material covered in this chapter in two 90-minute

lectures is available on the authors' websites.

2 Verification: The Permanent Income Hypothesis

Natural experiments can be used in macroeconomics to test the validity of major underlying model assumptions. This is done in the use of natural experiments to test the validity of the Permanent Income Hypothesis. The Permanent Income Hypothesis, as developed by [Friedman \(1957\)](#), contrasts with the simple Keynesian consumption theory, which postulates that consumption depends on current income only and is equal to a non-increasing fraction of current income. To the present day, the Permanent Income Hypothesis is the major building block of modern consumption theory, for example, the life cycle theory, the precautionary savings theory, and also behavioral consumption models involving hyperbolic discounting. The most important insight of the Permanent Income Hypothesis is that individuals are rational and forward looking when making their consumption decisions over the life cycle.

According to the Permanent Income Hypothesis, individual i solves a utility maximization problem of the form

$$\max_{\{C_{i,t+j}\}_{j=0}^{\infty}} E_t \sum_{j=0}^{\infty} \beta^j u(C_{i,t+j}) \quad (1)$$

subject to the intertemporal budget constraint

$$\sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j C_{i,t+j} = A_{i,t} + \sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j Y_{i,t+j}, \quad (2)$$

where $C_{i,t}$ is consumption of individual i in period t , β is the discount factor, r is the interest rate, $A_{i,t}$ are initial assets in period t , $Y_{i,t}$ is income in period t , and E_t is the expectations operator conditional on information available at time t . For simplicity, let us assume $\beta(1+r) = 1$. Also for simplicity, let's assume for now that the utility function takes the quadratic form, such that certainty equivalence holds:

$$u(C_{i,t+j}) = C_{i,t+j} - \frac{\alpha}{2} C_{i,t+j}^2. \quad (3)$$

This simple model has several powerful implications. Most importantly, consumption is not a function only of current income. Instead, it also depends on current assets and

expected future income, and is in fact equal to permanent income. Permanent income is defined as the annuity value of total net worth, which is the sum of current assets and the expected discounted net present value of all future income streams:

$$C_i = \frac{r}{1+r} \left[A_{i,t} + E_t \left(\sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j Y_{i,t+j} \right) \right]. \quad (4)$$

Because the expected discounted net present value of future income enters the optimal consumption decision of an individual, optimal consumption will change whenever new relevant information arrives. Conversely, any *anticipated* change in income will not affect optimal consumption. Consumption growth depends only on changes in the information set between periods t and $t+1$. Thus, we have

$$\Delta C_{i,t+1} = \frac{r}{1+r} \left[E_{t+1} \left(\sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j Y_{i,t+j+1} \right) - E_t \left(\sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j Y_{i,t+j+1} \right) \right] \quad (5)$$

and specifically

$$\Delta C_{i,t+1} = 0 \text{ if } E_{t+1} = E_t. \quad (6)$$

Equation (6) holds independent of the form of the utility function used in (1), as long as the desired consumption path is flat. The predictions from equations (5) and (6) can be tested by analyzing the reaction of consumption to anticipated and unanticipated income changes in the data. The empirical challenge lies in identifying in the data whether the individual anticipated any observed income change, and natural experiments are used to identify clearly unexpected or clearly anticipated income changes.

We start out describing the few papers analyzing the reaction of consumption to unexpected income shocks. The literature on the reaction of consumption to anticipated income changes is much larger, for reasons described below, and we will use this literature to gain more insights into the specifics of the use of natural experiments.

2.1 Reaction of Consumption to Unexpected Income Shocks

Only a few papers test whether consumption responds to unanticipated income shocks as predicted by equation (5). The reason is that the specific optimal reaction of consumption to an income shock depends among other things on the nature of the shock (whether it

is temporary or permanent), on the age of the recipient (if we deviate from an infinite horizon assumption and instead employ a life-cycle set up), and on the functional form of the utility function, which in a more realistic set up might involve prudence from part of the household, such that households build a buffer stock of savings to partly self-insure against future income fluctuations.

2.1.1 Unexpected Temporary Income Shocks

If we maintain the assumption of a quadratic utility function, and if an unexpected income change, that is, an income shock, is a strict one-time temporary income change, equation (5) reduces to

$$\Delta C_{i,t+1} = \frac{r}{1+r} [Y_{i,t+1} - E_t(Y_{i,t+1})]; \quad (7)$$

that is, the predicted consumption change is equal to the annuity value of the unexpected income change. Thus, as a generalization of this prediction, the predicted consumption change after a temporary income shock clearly should be small. One therefore needs large temporary income changes in the data in order to identify the response of consumption.

A very early paper testing this prediction is [Kreinin \(1961\)](#), whose analysis was later supported by further evidence by [Landsberger \(1966\)](#). [Kreinin \(1961\)](#) uses the 1957/58 Israeli Survey of Family Savings to analyze how Israeli households spent one-time restitution payments from Germany, which around 4 percent of urban Israeli households received during the year of the survey. He finds that Israeli households saved approximately 85 percent of the restitution payments, which on average amounted to close to one annual disposable income.¹ This behavior seems roughly in line with a small response of consumption to the temporary income change.

[Imbens et al. \(2001\)](#) and [Kuhn et al. \(2011\)](#) analyze the consumption of lottery winners. Lottery wins are historical episodes that clearly identify random large temporary income shocks, and can as such be seen as natural experiments. [Kuhn et al. \(2011\)](#) compare consumption of Dutch lottery winners and non-winners.² The lottery wins in their episode amount to 12,500 Euros, which is equal to eight monthly average household incomes in the Netherlands. In line with the Permanent Income Hypothesis, [Kuhn et al. \(2011\)](#) find that nondurable consumption does not increase significantly through a lottery win, but

¹By contrast, [Bodkin \(1959\)](#) finds that windfall incomes of National Service Life Insurance dividends paid out to US veterans were largely consumed. However, these windfalls amounted to, on average, only around five percent of annual disposable income.

²They also analyze social effects in a partial population design.

durable expenditures increase somewhat. [Imbens et al. \(2001\)](#) analyze significantly larger lottery wins, which are reimbursed over 20 years, and find that the increase in savings after a win is in line with the life cycle hypothesis. The authors of both studies collect their own data by sending out questionnaires to lottery winners and a sample of non-winners. The final sample sizes are then comparatively small, with 220 lottery winners in [Kuhn et al. \(2011\)](#), and 340 in [Imbens et al. \(2001\)](#).

[Brueckner and Gradstein \(2013\)](#) take a macroeconomic approach to analyze the response of consumption to unexpected temporary income shocks. Exploiting the fact that rainfall is a significant driver of annual aggregate output in sub-Saharan African countries, and that annual variations in rainfall are random and unexpected, they use rainfall as an instrument for aggregate output in a regression that analyzes the reaction of aggregate private consumption to aggregate output. They estimate a marginal propensity to consume out of temporary output shocks that is not significantly different from 0, with a point estimate of 0.2. Thus, similar to the studies relying on micro data, they find evidence of significant consumption smoothing of temporary income shocks.

2.1.2 An Unexpected Permanent Income Shock: the Natural Experiment of German Reunification

Germany's separation and subsequent reunification constitute in many ways a perfect natural experiment. A country with a common history is split into two parts and lives under very different economic and political systems for 40 years before being reunified. Importantly, it can confidently be argued that the separation of Germany was exogenous to the preferences of the underlying populations and the economic conditions in East and West at the time. The exact location of the border was largely determined by the position of the allied forces at the end of the war, which in turn was partly determined by the geographic location of the allies vis-a-vis Germany. To put it bluntly: if the Soviet Union would have been located to the West of Germany, some western part would have been socialist for 40 years. That the location of the East-West border can be considered random is best documented in the paper by [Alesina and Fuchs-Schündeln \(2007\)](#), who provide an overview of the economic and political situation in Germany before World War II, and show that no marked differences existed between East and West prior to separation. Based on this evidence, West Germans can be taken as a control group for East Germans, and economic conditions of East Germans at reunification, resulting from living under the socialist system of the former German Democratic Republic for

40 years, can be considered exogenous with respect to the new economic system after reunification, since German Reunification was an unexpected surprise event. This is a large-scale experiment, affecting close to 20 million people in a multitude of dimensions.

German Reunification has thus been used in a number of studies in the last two decades to analyze different questions. The first paper using German Reunification as a natural experiment is [Fuchs-Schündeln and Schündeln \(2005\)](#), who analyze self-selection in occupational choice according to risk preferences and its effects on estimates of precautionary savings. [Redding and Sturm \(2008\)](#) study the role of market access, and [Redding et al. \(2011\)](#) and [Ahlfeldt et al. \(2015\)](#) focus on industrial location choices. [Gebhardt \(2013\)](#) uses German reunification as a natural experiment to analyze the effect of ownership on relationship specific investment in the housing market, and [Burszтын and Cantoni \(2015\)](#) to investigate the effect of television advertisement on consumption. The studies by [Alesina and Fuchs-Schündeln \(2007\)](#), analyzing endogenous preferences for redistribution, and [Burchardi and Hassan \(2013\)](#), studying the effect of social ties on growth, are described below and also rely on the natural experiment of German reunification.

In the context of the Permanent Income Hypothesis, [Fuchs-Schündeln \(2008\)](#) exploits German Reunification as a large positive permanent income shock for East Germans. This permanent income shock is embedded into a structural life cycle model of consumption. This is one of the very few papers which combine a structural model in macroeconomics with a natural experiment.³ As in any structural model, this implies making assumptions about functional forms and calibrating the model carefully. Yet, it has the advantage that one can talk about the match between quantitative model implications and the data, and can distill the relative importance of different model components.

The life cycle model in [Fuchs-Schündeln \(2008\)](#) incorporates a retirement saving motive, a precautionary saving motive due to income risk and an exogenous liquidity constraint, and deterministically changing household size over the life cycle. West German life cycles play out in this model context from start to end, but East German households enter the new economic model environment in 1990 at a certain age. At this point in time, they are endowed with an exogenous wealth level, which is taken as the cohort-specific East-West wealth ratio in 1990 from the data. Importantly for the predictions of the model, the East-West wealth ratio at reunification was very low (lower than the East-West income ratio), which is especially true for older cohorts closer to retirement. From that point on, East Germans also live in this new economic model environment. Life-cycle

³This approach is more common in other fields, see e.g. the paper by [Ahlfeldt et al. \(2015\)](#).

income growth, income risk, and changing household sizes are calibrated separately for East and West Germans.

The calibrated model is able to qualitatively and quantitatively match three stylized features of East and West German saving rates after reunification: (i) East Germans have higher saving rates than West Germans; (ii) this East-West saving rate difference is increasing in age at reunification; that is, it is larger for older birth cohorts than for younger birth cohorts; and (iii) for every birth cohort, this difference is declining over time, with full convergence of saving rates within roughly ten years. The higher East German saving rates after reunification are a result of their low initial wealth levels, which leave them unprepared for the new economic environment in terms of both precautionary and retirement savings. The East-West difference in saving rates is especially large for older cohorts, because older cohorts of East Germans are least prepared for the new environment: their wealth position relative to their West German counterparts is especially low, and they have less time left over their working life to accumulate more wealth through higher saving rates. The rapid convergence of East German saving rates toward West German levels is the stylized feature that allows for differentiation between the different components of the life cycle model. A precautionary savings motive is essential to replicate this feature, because precautionary savings imply that saving rates decrease as wealth levels approach the target level of wealth from below.

The demographic developments after reunification alone would actually predict rising East German saving rates for younger cohorts, running counter to the empirical evidence. Disentangling a precautionary saving motive from a demographic saving motive based on changing household size over the life cycle is difficult in a standard setting, since both saving motives predict a hump-shaped consumption path over the life cycle. In the context of the natural experiment of German reunification, however, both saving motives lead to opposite predictions for the saving behavior of East Germans relative to West Germans. The paper concludes that East Germans react according to the predictions of the life cycle model after the large shock of German Reunification, despite being confronted with entirely new economic conditions, and that a precautionary saving motive is essential for replicating the data. The first conclusion is in line with the conclusions of the other studies analyzing large temporary income shocks. The second conclusion is only possible in a structural model, pointing to the advantages of the approach used in this paper. Relying on a structural model, one can go beyond analyzing main model predictions to analyzing the importance of different specific model components.

2.2 Reaction of Consumption to Expected Income Changes

In this section, we describe the literature using natural experiments to test the prediction of the Permanent Income Hypothesis that consumption growth should be insensitive to preannounced income changes, as specified in equation (6). This is a very large literature: Appendix table 1 discussed in section 2.2.3 below lists 24 published studies directly testing this prediction, and six further studies related to it in some way. We first focus on the methodological side by describing the use of natural experiments, then discussing in section 2.2.1 different ways to support the random treatment assumption in these studies, and next analyzing how the presence of liquidity constraints modifies the predictions of the theory, and how the papers deal with liquidity constraints. Section 2.2.3 then turns away from the methodology to focus on the findings of the studies, and section 2.2.4 tries to reconcile these sometimes contradictory findings by organizing them along two lines: the size of the income change and the repetitiveness of the episode under study.

The second implication of the Permanent Income Hypothesis - that an *anticipated* income change should not lead to a change in consumption - has the advantage of holding independently of the concrete set up of the problem. In particular, it holds also under functional forms of the utility function other than the quadratic one (e.g. under constant relative risk aversion), independent of the age of the individual in a life-cycle set up, and independent of the permanency of the income change at hand. This prediction can be tested if the econometrician knows that an observed income change was anticipated; that is, $Y_{t+1} \neq Y_t$, but $E_{t+1} = E_t$. The null hypothesis would then be that $\Delta C_{t+1} = 0$ and can be tested against the alternative $\Delta C_{t+1} \neq 0$ in a simple reduced-form regression of the form

$$\Delta C_{i,t+1} = \alpha + \beta \Delta Y_{i,t+1}^{expected} + \gamma' \Delta X_{i,t+1} + \epsilon_{i,t+1}, \quad (8)$$

where X is a vector containing any characteristics that are relevant for consumption and might have changed over time, for example, age and household size. The identifying assumption is that the error term is uncorrelated with the expected income change, that is, $Cov[\Delta Y_{i,t+1}^{expected}, \epsilon_{i,t+1}] = 0$, meaning no unobserved variables are correlated with the expected income change and the consumption change. The Permanent Income Hypothesis states that $\beta = 0$. If the underlying assumption of rational expectations and forward looking behavior is violated, we would expect that $\beta \neq 0$, and specifically that $\beta > 0$ under the Keynesian consumption theory.

Running this regression is easy if an expected income change can be directly observed

in the data, that is, if we know the underlying assumption $E_{t+1} = E_t$ holds. However, in general, whether any observed income change in the data was expected or unexpected by the individual is unclear. A common way to run this regression in the macro literature relying on aggregate consumption data involves the use of instruments. For example, [Ludvigson and Michaelides \(2001\)](#) regress quarterly consumption changes on quarterly income changes, instrumenting income changes with their own lags. [Carroll and Summers \(1991\)](#) run similar regressions on international data, again instrumenting with lags of income growth. However, at the micro level, to which the theory applies, finding a suitable instrument is much harder.

A more elegant and convincing way to run this regression on the micro level is to exploit a natural experiment. Natural experiments in this context are clear historical episodes in which we know that an income change occurred, and that it was preannounced and thus anticipated by the households. Typical income changes of this kind analyzed in the literature are associated with taxation (tax rebates, tax refunds, changes in tax laws, etc.), wages (wage payment schedule, wage changes, social security receipts), and committed consumption (college cost, mortgage payments, etc.). All these changes have in common that they are clearly announced some time in advance, and thus the recipient anticipates them. The Permanent Income Hypothesis predicts that households should adjust their consumption at announcement of the income change. The size of the optimal consumption adjustment at announcement depends among other things on the expectations about the exact nature of the income change and is therefore hard to gauge, as in the papers described in section 2.1. By contrast, testing the prediction that consumption should not react when the preannounced income change actually happens is easy.

In a more general sense, one can think of the test for whether $\beta = 0$ in equation (8) as a general test of the validity of the rational expectation assumption in consumption decisions. We might not care from either a macro or micro point of view whether households adjust their consumption at the announcement or the implementation of an income change, because both typically happen within a short period of time in the natural experiments analyzed in the literature. However, for welfare purposes, whether households build rational expectations and are forward-looking when deciding how much to consume and how much to save matters tremendously. For example, to save appropriately for retirement, households have to understand the income process over their life cycle early on and act accordingly.

2.2.1 Random Treatment: Determining an Appropriate Control Group

The estimation of equation (8) using a natural experiment to establish that an income change was anticipated still faces some challenges. Importantly, equation (8) can only be estimated consistently if the error term is uncorrelated with the preannounced income change; that is, $Cov[\Delta Y_{t+1}^{expected}, \epsilon_{t+1}] = 0$. Otherwise, the preannounced income change and the consumption change would be spuriously correlated due to omitted variables.

One important feature that could lead to correlation between the error term and the preannounced income change could be seasonality effects. For example, workers in many countries receive a 13th salary in the month of December, leading to a preannounced change in monthly income between November and December. At the same time, expenditures increase in December because of holiday shopping. This leads to a spurious correlation between the preannounced income change and the consumption change. The income change is endogenous because the 13th salary in December was established precisely because firms recognized the higher average household expenditure in December.

In the spirit of an experimental set up, a valid control group can overcome this problem. If the above-mentioned preannounced income change exhibits temporal variation, that is, if it does not occur in the same month for all households, then variation is present in the timing of the treatment, and one can include monthly dummies to account for seasonality in expenditures directly. The same applies if the preannounced income change happens in different months in different years, though in that case, one has to argue that expenditure seasonality should be the same year by year, for example by analyzing whether major events usually causing increases in expenditure, like public holidays or vacations, happen in the same months every year. Variation in the individual amount of the preannounced income change relative to permanent income could help, but only if one could reasonably argue that this variation is exogenous to any desired seasonality in expenditure.

In the ideal experiment, one group does not receive any preannounced income change, and another one does, and both should be comparable along all other observable and unobservable characteristics, including preferences that lead to consumption seasonality. In that case, one can think of the first group as the “control” group and of the second group as the “treatment” group. Here, the natural experiment is very close to a designed field or laboratory experiment: two groups exist, one of which is quasi-randomly treated and the other one not, and the behavior of both groups is compared. The analysis of consumption changes then corresponds to a difference-in-differences set-up. Whereas laboratory or field experiments would be designed to make the assignment into the treatment group

explicitly random, the main challenge of a natural experiment is to convincingly argue the randomness of the assignment and thus the appropriateness of the control group. Arguing this point is generally easiest if both groups receive the same treatment, but at randomly different points in time. This distinguishes natural experiments from field or laboratory experiments, which typically leave a control group untreated.⁴

In this section, we describe different methods to determine randomness in treatment. In passing, we also discuss some findings of the papers, which are, however, the focus of section 2.2.3.

Clearly Established Randomness in Treatment A set of studies that are particularly successful in establishing randomness in the treatment assignment are the papers by [Johnson et al. \(2006\)](#) and [Agarwal et al. \(2007\)](#), who exploit the 2001 Federal Income Tax Rebates as a natural experiment, and the studies by [Parker et al. \(2013\)](#) and others, who analyze the 2008 Economic Stimulus Payments as a natural experiment.⁵ In both cases, households received one-time tax rebates in the form of checks sent to them. The media extensively discussed the rebates in advance, and as such, households should have known about them. In addition, for the 2001 Bush tax rebates, households received an individual letter several months in advance stating the specific amount of the rebate.⁶ Although the amount received varied little between households, mostly driven by household size and thus not exogenous, nice and clearly exogenous variation exists in the timing of the payments: because sending out all rebate checks on the same day was logistically impossible, the IRS spread out the payments over ten weeks in 2001 and nine weeks in 2008, and determined the exact date on which each household would receive the check by the second to last digit of the Social Security Number of the main tax payer, which is randomly assigned. Thus, in these two cases the randomness in the timing of treatment is as clearly established as in any field or laboratory experiment in which the researcher consciously

⁴A valid reason for this approach for field or laboratory experiments is the fact that treatment is typically costly for the researcher.

⁵[Johnson et al. \(2006\)](#) and [Parker et al. \(2013\)](#) analyze consumption responses, whereas [Agarwal et al. \(2007\)](#) analyze the response of credit card spending and debt repayment to the 2001 federal income tax rebates. The 2008 Economic Stimulus Payments have been exploited by a number of studies, including [Broda and Parker \(2014\)](#) and [Parker \(2014\)](#) analyzing consumption responses, [Gross et al. \(2014\)](#) and [Bertrand and Morse \(2009\)](#) analyzing bankruptcy filing and repayment of payday loans, respectively, and [Evans and Moore \(2011\)](#) and [Gross and Tobacman \(2014\)](#) analyzing mortality and morbidity outcomes. [Shapiro and Slemrod \(2003\)](#) and a series of papers by [Sahm et al. \(2009, 2010, 2012\)](#) analyze self-reported propensities to consume and to save out of both rebate episodes. [Misra and Surico \(2014\)](#) analyze heterogeneity in consumption responses to both the 2001 and 2008 stimulus programs.

⁶For the 2008 Economic Stimulus Payment, the letter came only one week in advance.

establishes randomness through a lottery. Exploiting this situation, the “treated” group in the above-mentioned studies is the one that randomly receives the rebate in the period under consideration, whereas the “control” group encompasses all other households, which receive the rebate in a different period.⁷ The week of rebate receipt is clearly exogenously determined. All of these studies find that household consumption adjusts at receipt of the rebates, in violation of the Permanent Income Hypothesis.

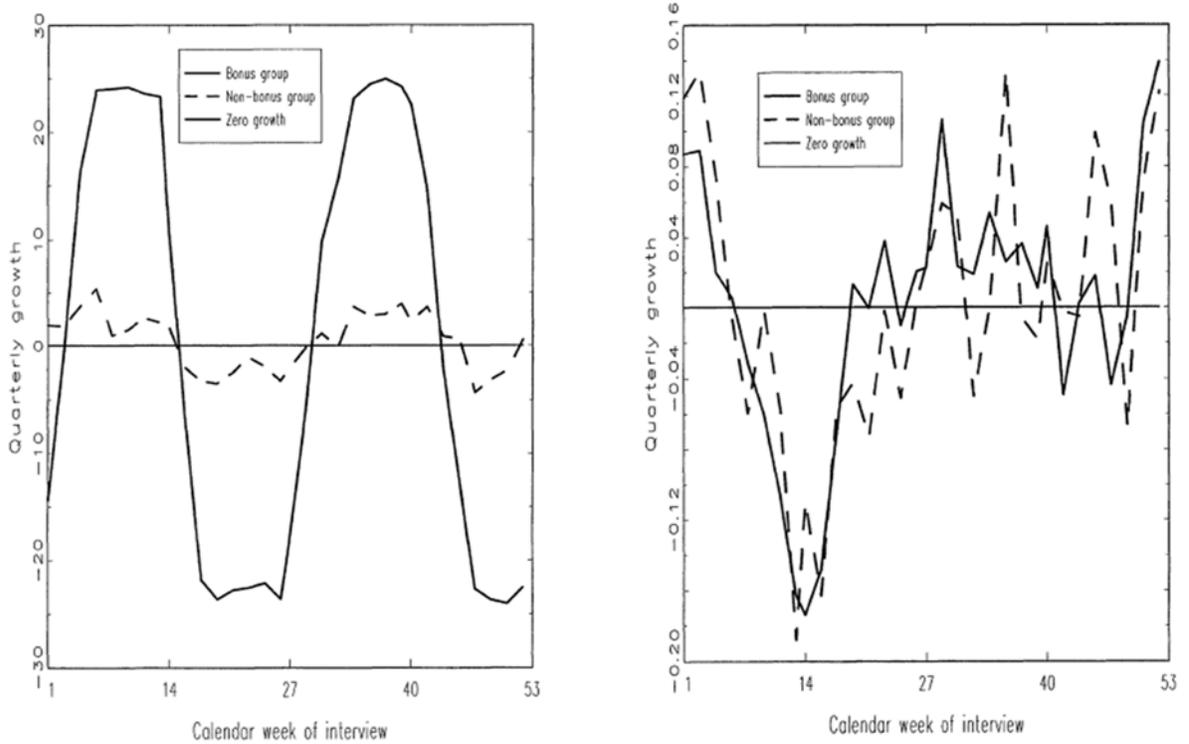
The “Narrative Approach” In the absence of such a clear random treatment assignment, different strategies can be used. For example, [Browning and Collado \(2001\)](#) analyze quarterly consumption growth of Spanish workers who are part of one of two different payment schemes: the standard scheme used in the control group encompasses monthly wage payments of twelve equal amounts over the year, whereas the second payment scheme in the treatment group involves higher payments in the months of June (or July) and December. The payment scheme varies on the plant level, and because workers know which payment scheme their plant follows, workers in the treatment group should perfectly anticipate the unusually high monthly income growth between the months of May and June (or June and July, if the extra payment is in July rather than June) as well as November and December, each followed by a month of unusually low income growth. To test the prediction of the Permanent Income Hypothesis that consumption growth should not react to preannounced income changes, the authors then simply compare seasonal consumption patterns of the treatment group (called “bonus group”) and the control group (called “non-bonus group”).⁸ [Figure 1](#) shows quarterly income growth per calendar week of both groups on the left, and quarterly expenditure growth on the right.⁹ Despite strong differences in the income growth patterns between both groups, the expenditure growth patterns are very similar. Thus, the evidence in [Browning and Collado \(2001\)](#) is in line with the predictions from the Permanent Income Hypothesis. The major challenge here is to argue about the random assignment of the payment scheme. For example, plants might use the second payment scheme because they know their workers have unusually strong

⁷In the case of the 2008 Economic Stimulus Payments, part of the households received not a check but a direct deposit, for which the timing was somewhat different. Thus, the studies using the 2008 Economic Stimulus Payments suffer from larger measurement error than the studies using the 2001 federal income tax rebates, if they cannot determine whether a household received a check by mail or a direct deposit, which most cannot.

⁸Similarly, [Hsieh \(2003\)](#) compares the seasonal consumption patterns of Alaskans to the seasonal consumption patterns in other US states, and [Paxson \(1993\)](#) compares seasonal consumption patterns of farmers and non-farmers in Thailand.

⁹Income is measured as average income in the three quarters preceding the interview. While the extra payments are called “bonus”, there is no performance component involved.

Figure 1: Quarterly earnings growth (left) and quarterly total expenditure growth (right) in Browning and Collado (2001)



preferences for seasonally high expenditures in June/July and December, for example, due to certain holiday traditions in their region. The authors explicitly discuss this assumption and give some historical account of how the two payment schemes arose. They also show that being part of either payment scheme is not significantly correlated with any observable household characteristics. We call this the “narrative approach”, because it relies purely on carefully arguing about exogeneity of the treatment, and ruling out potential alternative stories of endogeneity. Placebo exercises, described below, are useful in this regard. Since Browning and Collado (2001) find that expenditure growth patterns of both groups over the year resemble each other, the argument about exogenous treatment becomes somewhat less important; any endogeneity should have led to the observation of stronger seasonal expenditure patterns correlated with the preannounced income changes for the treatment group.

Using Different Control Groups and the Matching Approach Sometimes doubts about exogeneity of the treatment remain even after a detailed description and careful

analysis of the circumstances leading to treatment versus non-treatment in the “narrative approach”. In this case, one can follow different strategies to still establish some confidence into a causal effect. The most basic strategy, followed by many papers, is to establish robustness of the results to the use of different control groups. Consider, for example, the study by [Agarwal and Qian \(2014a\)](#), who analyze the response of consumption and debt repayment to a unique cash pay-out by the government to each adult Singaporean. The pay-out happened at the same time for all eligible individuals, such that no randomness in the timing was present. Although amounts varied across individuals, this variation was not random, because the amount was a function of income and home values. [Agarwal and Qian \(2014a\)](#) use foreigners living in Singapore as a control group: foreigners make up almost 40 percent of the population living in Singapore and were not eligible to receive the pay-out. They show results of their analysis using this control group, as well as restricting the analysis to Singaporeans and exploiting only the (non-random) variation in amounts.

Both approaches clearly have their disadvantages. Specifically, foreigners only constitute a valid control group if their spending patterns are similar to those of Singaporeans in the absence of treatment. In a first step, the authors compare Singaporeans and foreigners along observable characteristics and find some significant differences. To control for these observable differences, they use propensity score matching methods (going back to [Rosenbaum and Rubin \(1983\)](#)) to construct two subsamples of matched treatment and control groups that are comparable across most observable characteristics. Researchers frequently use propensity score matching methods in microeconomic set ups in which random treatment cannot be assumed. The basic idea behind a variety of sub-methods is that one constructs a subsample of the treatment group and a subsample of the control group, which are as comparable as possible along a long list of observable characteristics. Importantly, [Agarwal and Qian \(2014a\)](#) also show that both subsamples have comparable seasonal spending patterns prior to the treatment, though this information is not used to construct the subsamples. [Agarwal and Qian \(2014a\)](#) find that Singaporeans increase their consumption already at announcement of the pay-out, and spread the consumption increase over the following 10-month period.

[Abdallah and Lastrapes \(2012\)](#) use a similar approach, analyzing the effect of a pre-announced relaxation in the borrowing constraint among Texan home owners in 1997 on Texan retail spending. They start out using two control groups, the first consisting of all other US states except Texas, and the second consisting of all other US states that did not change sales tax rates during the estimation period. They allow for state-specific linear time trends, in order to ensure that a different general time trend in Texan retail spend-

ing is not mistakenly attributed to the policy change. In a next step, they also employ a form of matching methodology, specifically, the synthetic control method of [Abadie and Gardeazabal \(2003\)](#) and [Abadie et al. \(2010\)](#), which assigns optimally selected weights to each control group observation in order to minimize the distance between predicted sales in Texas and the control group during the pretreatment period. This study falls into the group of studies analyzing liquidity constraint relaxations (see the discussion in section 2.2.3 below), and like all these studies, finds evidence for binding liquidity constraints.

The Use of Placebo Exercises A formal way to gauge the validity of the control group is the use of placebo exercises. The idea of this approach is to define virtual “placebo treatments” and to compare the average effects of these “placebo treatments” to the one of the actual treatment. Consider again the study by [Abdallah and Lastrapes \(2012\)](#). In this study, the “placebo treatment” can be defined as dropping Texas from the analysis and assuming a state other than Texas introduced a similar relaxation of the borrowing constraint at the same point in time when Texas actually did. If one includes all US states in the analysis, one ends up with 49 different “placebo treatments.”¹⁰ For each of them, the baseline regression is run, and the baseline estimate on the true treatment is compared to the distribution of the estimated coefficients β from the placebo treatments. The treatment effect from the baseline regression should be well above the median placebo treatment effect in order to confirm a true effect. This approach can also be applied to control groups built based on a matching method. [Gross et al. \(2014\)](#), [Mastrobuoni and Weinberg \(2009\)](#), and [Scholnick \(2013\)](#) perform similar placebo exercises.

2.2.2 The Presence of Liquidity Constraints

As stated above, equation (6) holds under different concrete setups, for example, in an infinite or a finite life-time setting, under different assumptions of the functional form of the utility function, and so on. However, one important assumption has to be maintained: the consumer problem laid out in equations (1) and (2) does not contain a liquidity constraint. If liquidity constraints are present and binding, the household will not be able to adjust consumption optimally at the announcement of a future income increase, but only at the implementation of the income increase.

We can deal with this complication in two ways. First, and most convincingly, one can analyze the consumption reaction to a preannounced income *decrease* rather than an

¹⁰In addition, one could specify “placebo treatments” taking place in Texas, but at a different point in time, or even taking place in other states at a different point in time.

increase. The presence of a liquidity constraint does not affect the optimal consumption change triggered by the announcement of a future income decrease: decreasing consumption is always a possibility. Unfortunately, the vast majority of the “natural” situations that researchers can analyze involve income increases rather than decreases. Here, the limitation of natural experiments, which cannot be designed to prevent certain limitations a priori, becomes clear in contrast to self-designed field or laboratory experiments. One paper that does analyze a decrease in income is the study by [Souleles \(2000\)](#). The paper analyzes the change in consumption upon an increase in college expenditure due to a child in the household starting college. Because the college entrance can be foreseen for some time, and college costs are also usually determined in the spring before the start of college, one can think of the increase in college expenditure as a perfectly anticipated increase in committed consumption and therefore as a decrease in net disposable income.¹¹ [Souleles \(2000\)](#) finds that expenditure on strictly nondurable goods and food does not fall significantly upon the anticipated decrease in net disposable income, or if anything, it falls by a very small amount, depending on the specification. This study thus finds behavior in line with the Permanent Income Hypothesis if households face an anticipated income decrease.

Apart from focusing on income decreases, one can explicitly analyze the importance of liquidity constraints by splitting the sample into potentially constrained and most likely unconstrained households. Because binding liquidity constraints are typically unobservable, this analysis is approximate rather than exact. Consider, for example, the paper by [Johnson et al. \(2006\)](#) cited above. They rely on three different measures that can proxy for liquidity constraints: age, income, and liquid assets, where the assumption is that young households, households with low income and/or a low levels of liquid assets are more likely to be liquidity constrained. Splitting the sample into “high,” “medium,” and “low” values of the respective variable, they then analyze whether the consumption change of the “low” and thus potentially liquidity constrained group is larger than that of the other groups.¹² Moreover, the coefficient β should be equal to zero for the unconstrained “high” group in the absence of measurement error in measuring liquidity constraints. The evidence points towards liquidity constraints: the “low” group increases consumption more upon

¹¹This decrease in net disposable income is of course endogenous, because parents can choose whether and how much to spend on a child’s college education. The paper includes robustness checks instrumenting for college expenditure.

¹²Part of the literature relying on credit card data can apply more direct measures of liquidity constraints, for example, whether individuals regularly pay interest on their credit card, or how close they are to the credit limits (see [Agarwal et al. \(2007\)](#), [Agarwal and Qian \(2014a\)](#), and [Scholnick \(2013\)](#)).

receipt of the rebate check than the “high” group, though the difference is not always statistically significant. Moreover, even the “high” group shows a positive consumption response under some measures. These are two common findings in the literature: potentially liquidity constrained groups react stronger upon payment receipt, but groups that are likely not liquidity constrained still react significantly.¹³ Misra and Surico (2014) point out that traditional measures of liquidity constraints might miss wealthy hand-to-mouth consumers (Kaplan and Violante (2014a)), who hold most of their assets in illiquid form, and provide evidence that indeed these consumers also react strongly to preannounced income changes. They argue that imposing homogeneous consumption responses not only misses interesting and systematic variation in the data that is potentially linked to liquidity constraints, but can also lead to biased estimates of the average consumption response. Indeed, analyzing measures of liquidity constraints that also incorporate wealthy hand-to-mouth consumers might allow reconciling some of the conflicting evidence on liquidity constraints found so far.

Another set of studies analyzes consumption during the payment cycle. Because income increases from zero to a positive value on the day of wage payment receipt, and then falls to zero again the day after payment receipt, these studies also encompass regular income decreases. These studies typically involve frequencies higher than the monthly one and analyze whether consumption is stable or decreases over the payment cycle (see Gelman et al. (2014), Mastrobuoni and Weinberg (2009), Shapiro (2005), Stephens (2006), and Stephens (2003)). In principle, if one assumes the first payment comes at the end of a “consumption cycle,” liquidity constraints could matter. However, if these regular payments have already been received for some time, arguing that households could not build up a buffer to smooth variations over the pay cycle is hard. Still, some of these studies employ explicit tests for liquidity constraints as described above and find evidence in favor of liquidity constraints (Gelman et al. (2014), Mastrobuoni and Weinberg (2009), and Stephens (2006)). Because, in principle, liquidity constraints should not matter in this setup, the evidence in favor of liquidity constraints might indicate that the proxies for liquidity constraints are correlated with other behavioral traits that could drive the excess sensitivity of consumption to preannounced income changes.

¹³Unfortunately, not all studies analyzing liquidity constraints show results testing the latter hypothesis that non-liquidity constrained groups should not react significantly.

2.2.3 Overview of Natural Experiment Studies of the Permanent Income Hypothesis

More than two dozen papers test equation (8) in various ways. Describing these papers in detail is beyond the scope of this chapter. Nevertheless, appendix table 1 at the end of the chapter provides a brief overview listing papers in alphabetical order. The table lists the nature of the specific episode that is analyzed and whether it involves an income increase or decrease, the data source (including country and specific data set used) and sample selection, the main dependent variable and its frequency, whether the paper finds significant evidence against the Permanent Income Hypothesis and what the main result is quantitatively, and finally whether any tests of liquidity constraints are carried out and what their results are. Unfortunately, because the concrete estimations run in each paper are different, one cannot indicate a comparable estimated coefficient β for each study, but we provide the main quantitative result as stated in the respective paper. All but two of the papers involve the use of household or individual data.

Appendix table 1 distinguishes between three different sets of studies. The first set includes 25 studies that analyze an experiment involving an anticipated change in disposable income, most often because of a direct gross or net income change, sometimes because of a change in payment commitments from mortgages or college expenditures. The dependent variable in these studies is some measure of consumption, which varies from standard measures of non-durable or durable consumption over caloric intake to credit card spending or retail sales. Most of these studies find evidence against the Permanent Income Hypothesis: consumption growth reacts to the implementation of the preannounced income change. Notable exceptions are the studies by [Agarwal and Qian \(2014a\)](#), [Browning and Collado \(2001\)](#), [Coulibaly and Li \(2006\)](#), [Hsieh \(2003\)](#), [Paxson \(1993\)](#), and [Souleles \(2000\)](#).

The second set includes four studies that analyze the reaction to preannounced relaxations of borrowing constraints. These studies can be seen as direct tests of the presence of binding liquidity constraints: if liquidity constraints are not binding, then any preannounced relaxation of a constraint might lead to consumption reactions at announcement, but not at implementation.¹⁴ If liquidity constraints are, however, currently binding, then any relaxation of the constraint should lead to increased consumption at implementation. The studies either involve an experiment directly relaxing the borrowing constraint and

¹⁴Consumption might still react at announcement, because the possibility of liquidity constraints becoming potentially binding in the future affects consumption today.

then use a measure of consumption as the dependent variable, or they involve an experiment relying on an expected income change as in the first set of studies but use a measure of loan take-up or bankruptcy filing as the dependent variable. Some of these studies still analyze whether potentially liquidity constrained households react stronger upon implementation than households who are less likely to be liquidity constrained. All of these studies find that liquidity constraints do matter.¹⁵

The third set includes two studies that deal with experiments that involve temporary price cuts. Under the assumption of forward-looking behavior, expenditures on goods subject to a temporary price cut increase during the period of the price cut, but at the same time, expenditures on these goods decrease before or after the period of the price cut if goods exhibit some degree of durability and the period of the price cut is relatively short.¹⁶ These two studies find different results: Sales tax holidays seem to have long-lasting effects on purchases of some affected goods (Agarwal et al. (2013)), whereas the 2009 “Cash for Clunkers” program merely shifted the purchases of new cars in time (Mian and Sufi (2012)).

2.2.4 Violation of Rational Expectations or Need for Model Extension?

A vast majority of the natural experiments investigating the Permanent Income Hypothesis find evidence against it by rejecting the null that $\beta = 0$. This seems to indicate that households are in fact not forward looking when making their consumption decisions, even if they have to take into account only income changes that will occur a few months ahead. How can one then assume that households look many years ahead, as required for retirement planning, saving for childrens’ college expenditures, and so on? Thus, one of the major assumptions of the Permanent Income Hypothesis seems to be undermined. The evidence summarized in section 2.2.2 suggests liquidity constraints can help reconcile theory and evidence, but often a significant consumption response to preannounced income changes can be found even among likely unconstrained households. Apart from analyzing liquidity constraints, the data are rarely rich enough to provide further insights into the sources of the failure of the Permanent Income Hypothesis.¹⁷

¹⁵DeFusco (2014) and Agarwal and Qian (2014b) are two recent papers analyzing similar experiments that involve a relaxation of borrowing constraints for home owners in the first case, and an unexpected tightening in the second case.

¹⁶Expenditures on other goods might also change, depending on the degree of substitutability or com-

Table 1: Studies of the Permanent Income Hypothesis Sorted by Size and Regularity of the Income Change

	Small	Large
Regular	Aaronson, Agarwal, and French (2012) 0.03% Parker (1999) ^a 0.00038 % Parker (1999) ^b 0.82% Shea (1995) 0.0009%	Browning and Collado (2001) 2.61% Hsieh (2003) 4.79% Paxson (1993) - Souleles (1999) 1.24% Souleles (2000) ^c 5.24%
Irregular	Agarwal, Liu, and Souleles (2007) 0.22% Agarwal and Qian (2014) 0.04% Broda and Parker (2014) 0.31% Coulbaly and Li (2006) 0.56% Johnson, Parker, and Souleles (2006) 0.10% Parker, Souleles, Johnson, and McClelland (2013) 0.46% Scholnick (2013) 0.45% Souleles (2002) 0.01% Stephens (2008) 0.35%	

Note: Papers written in bold fail to reject the Permanent Income Hypothesis. The number after each study indicates the equivalent variation associated with the respective experiment. The equivalent variation is calculated as described in the text. The online appendix provides details on the calculation of the equivalent variation for each paper.

^aChange in social security tax rate

^bCap in social security withholding

^cBecause of the absence of suitable expenditure and income data, the equivalent variation is calculated with price-adjusted average quarterly spending from Johnson, Parker, and Souleles (2006).

Table 1 distinguishes the existing natural experiment studies along two lines: how *large* the analyzed income change is, and whether it happens *regularly* over the life cycle. We consider an income change as *regular* if it is a repeated phenomenon that occurs to an individual several times over the life cycle, for example, tax refunds or payment schemes that double the income every year in July and December. On the other hand, unique government interventions such as tax rebates due to a fiscal stimulus program, or the running out of mortgage or college payments are considered irregular events. To classify an episode as large or small, we resort to the equivalent variation as a measure of the welfare loss associated with a certain behavior.¹⁸ Specifically, we compare two hypothetical consumers over the course of one year only, considering monthly consumption,¹⁹ and assuming additive separability of monthly utility and no discounting: the first “rational” consumer smoothes a pre-announced income change x over the course of one year. This behavior is obviously not optimal, because optimality would call for smoothing the income change over the entire life cycle, but it is a good approximation for those regular income changes that occur once a year, and otherwise provides a lower bound of the welfare losses. We calculate the utility of this consumer over a year as $U^{rational}(c) = 12 * u(y + \frac{x}{12})$, where y is regular monthly consumption and x is the extra amount received in the experiment. The second “hand-to-mouth” consumer has the same baseline income as the “rational” consumer, but consumes the extra amount x analyzed in the specific episode entirely in the month of receipt rather than spreading it evenly over 12 months; that is, her utility is $U^{hand-to-mouth}(c) = 11 * u(y) + 1 * u(y + x)$. We then calculate the equivalent variation as the monthly consumption amount z we would have to add to the consumption of the “hand-to-mouth” consumer to make her as well off as the “rational” consumer, expressed as a percentage of regular monthly consumption.²⁰ In these calculations, we assume a constant relative risk aversion utility function with a risk aversion parameter of 2. We consider an experiment as *large* if the equivalent variation amounts to more than 1 per-

plementarity between goods.

¹⁷The recent study by [Parker \(2014\)](#) is a step in the right direction. It analyzes the spending response to the 2008 Economic Stimulus Payments using data from the Nielsen Consumer Panel, and augments these data with questionnaires that allow the author to draw conclusions about certain personal characteristics such as lack of planning, impatience, and inattention.

¹⁸The distinction in small and large shocks has already been suggested by, among others, [Browning and Collado \(2001\)](#), [Hsieh \(2003\)](#), and [Jappelli and Pistaferri \(2010\)](#).

¹⁹The time unit is a month rather than a year because most experiments use monthly data and involve an episode of a predicted income change in a specific month, that is, most papers in the literature follow this timing logic.

²⁰In other words, z solves $11 * u[y + z] + 1 * u[y + x + z] = 12 * u[y + \frac{x}{12}]$, and the equivalent variation is calculated as $EV = z/y$.

cent. The spirit of this exercise and the specific threshold of 1 percent are in line with the study done by [Chetty \(2012\)](#), who analyzes bounds on labor supply elasticities, allowing for adjustment costs or inattention resulting in households not reacting to tax changes, as long as the associated utility loss amounts to less than 1 percent in a life-cycle setup. The online appendix explains in detail which values are used to calculate the equivalent variation for each study.

In table 1, papers that fail to reject the Permanent Income Hypothesis are written in bold. As the table shows, four of the six studies that do not reject the Permanent Income Hypothesis analyze large income changes, three of these analyzing income changes that occur repeatedly over the life cycle, and one analyzing an irregular income change.²¹ Among the studies analyzing large income changes, only one ([Souleles \(1999\)](#), who analyzes tax refunds) rejects the Permanent Income Hypothesis. This study involves an experiment associated with an equivalent variation barely exceeding 1 percent, being the smallest among the “large” studies.

The remaining two studies that find support for the Permanent Income Hypothesis analyze small, irregular changes. [Coulibaly and Li \(2006\)](#) find that home owners smooth consumption over their last mortgage payment, after which disposable income increases. The episode is characterized as small, because the last mortgage payment is typically not high and these households are relatively well off; from a life-cycle perspective, mortgage payments, however, constitute substantial consumption commitments and thus reductions in disposable income. The last study, by [Agarwal and Qian \(2014a\)](#), analyzes the 2011 Singaporean Growth Dividends, which amounted to around 500 USD per individual. This study is different from the others in that it explicitly analyzes the consumption reaction at implementation *and* announcement, and finds that consumption increases already at announcement, but remains higher for almost one year.

Taken together, the evidence appears to imply that households tend to behave consistently with the Permanent Income Hypothesis when the stakes are high, that is, when dealing with large or repeated changes in their income. A simple way to rationalize the different results of the papers may be to consider models that allow for monetary or psychological adjustment costs of reoptimization. Moreover, learning on the part of the consumer might play a role. Monetary or psychological adjustment costs would point to near-rationality, as defined, for example, by [Cochrane \(1989\)](#). The evidence in favor of the

²¹[Paxson \(1993\)](#) does not provide enough information to calculate the equivalent variation as in the other studies. Still, it is clear that the utility loss for farmers not smoothing income fluctuations over the year would be large in the sense of an equivalent variation exceeding 1 percent.

Permanent Income Hypothesis coming from the studies analyzing large income changes is in line with the natural experiment studies analyzing income shocks in section 2.1, which all look at large shocks to income and do not find evidence against the Permanent Income Hypothesis. Moreover, in the two studies that analyze temporary price cuts, the study analyzing large price cuts (Mian and Sufi (2012)) finds evidence in favor of rational behavior, whereas the paper analyzing relatively small price cuts (Agarwal et al. (2013)) finds evidence against rationality. Near-rationality is in the spirit of the concept of inattentive consumers as in Reis (2006) or of inaction inertia as discussed in, for example, Anderson (2003).²² In Reis (2006), households with high planning costs become inattentive savers, which live according to a saving plan and let consumption absorb all income changes that are not large enough to trigger a reoptimization. Small income changes might fall into this category, and thus an inattentive saver would not adjust his or her consumption at arrival of new information on a future small income change, but rather consume the extra income at arrival. The announcement of a large future income change would instead trigger reoptimization.²³

Perhaps the most convincing evidence in favor of adjustment costs or near rationality comes from Hsieh (2003). Hsieh (2003) uses data on Alaskan households from the consumer expenditure survey in order to analyze two natural experiments on the same set of households. The first experiment involves tax refunds also analyzed by Souleles (1999). These refunds are anticipated, because the taxpayer has to calculate them when filing the tax return. Around three quarters of all taxpayers receive refunds, and the average refund on the household level amounts to 700 USD to 850 USD (in 1982-1984 USD, see Souleles (1999)). Hsieh (2003) finds that Alaskan households receiving a tax refund consume 28 percent of it in the quarter of receipt. He then runs a regression on the same set of households, in which he analyzes a different preannounced income change, namely, payments from Alaska's Permanent Fund. The Alaska Permanent Fund redistributes receipts from oil royalties as dividend payments to residents of Alaska. The amount of the payment has been increasing over time and varies between around 300 USD per person in 1984 and almost 2000 USD per person in 2000. Because every resident of

²²Early studies of near-rationality include Akerlof and Yellen (1985) and Mankiw (1985). For a recent study, see Hassan and Mertens (2014).

²³The survey responses analyzed by Shapiro and Slemrod (2003) do not, however, support the implications of models of inattentive savers or inattentive consumers: in contrast to the model's prediction, individuals who report to target spending are more likely to spend the 2001 tax rebates than individuals who don't target spending, whereas individuals who report to target saving show the same propensity to save the rebate as those who don't. Also, the survey evidence by Parker (2014) does not point towards inattentiveness.

Alaska, regardless of income and age, is entitled to this payment, the average household payment is quite high, substantially higher than the average tax refund. [Hsieh \(2003\)](#) finds that the same households that show significant excess sensitivity of consumption to the preannounced income changes caused by tax refunds do not show such excess sensitivity to the preannounced income changes caused by dividends from the Alaska Permanent Fund. This is strong evidence that the size of the welfare cost associated with failing to smooth the income change in question matters, because it comes from exactly the same set of households. [Scholnick \(2013\)](#) also reports direct evidence that the magnitude of the analyzed income change matters. He analyzes the reaction of credit card spending to the predictable changes in disposable income resulting from a household's final mortgage payment. Because the mortgage payment amounts vary by households, also relative to their income, he can analyze whether the size of the income change matters.²⁴ Indeed, he finds a positive reaction of credit card spending to the income increase after the final mortgage payment, in violation to the prediction of the Permanent Income Hypothesis, but the larger the preannounced income change is, the smaller the reaction.

Summarizing, the literature on the Permanent Income Hypothesis finds that liquidity constraints clearly matter for some households. For households that are not constrained, near-rationality is a likely candidate to explain their excess sensitivity to small anticipated income changes. Faced with large income changes, households seem to react in line with the Permanent Income Hypothesis and are thus forward-looking when making their consumption decisions.

3 Quantification: The Fiscal Multiplier

The size of the fiscal multiplier is a highly controversial topic in macroeconomics. The fiscal multiplier measures the size of the output change associated with a change in a fiscal instrument; that is,

$$\Delta Y_{t+1} = \alpha + \beta \Delta F_{t+1} + \gamma' \Delta X_{t+1} + \epsilon_{t+1}, \quad (9)$$

where β is the fiscal multiplier, Y is a measure of output, F is a measure of the fiscal instrument, and X is a vector of potential control variables, typically including lagged growth measures.

²⁴One might, however, be worried whether the variation in size is endogenous, and how this endogeneity would affect the estimates. In a recent working paper, [Kueng \(2015\)](#) similarly finds evidence for near-rationality by exploiting variation in the relative size of the income change across households relying on the same experiment as [Hsieh \(2003\)](#).

Although, in principle, running this regression with a time series of macroeconomic data is easy, the macroeconomic literature on fiscal multipliers faces one serious challenge: the change in government spending must be exogenous to economic growth. A standard measure of total government spending is certainly subject to reverse causality, such that the assumption $Cov[\Delta F_{t+1}, \epsilon_{t+1}] = 0$ does not hold. For example, automatic stabilizers such as medicaid and unemployment insurance lead to an increase in fiscal spending precisely when output growth is low, such that $Cov[\Delta F_{t+1}, \epsilon_{t+1}] < 0$, biasing the estimate of β towards zero. On the other hand, procyclical government spending components might exist if governments have limited ability to accumulate debt, in which case $Cov[\Delta F_{t+1}, \epsilon_{t+1}] > 0$, and the estimate of β would be biased upwards.

The macroeconomic literature typically addresses the issue of endogeneity using vector auto regression (VAR) methods imposing identifying restrictions (see, e.g., [Blanchard and Perotti \(2002\)](#) and [Mountford and Uhlig \(2009\)](#)), for example, that government spending does not react to current economic conditions at the quarterly frequency, or relying on the so-called narrative approach, which establishes exogeneity of fiscal policies to current economic conditions based on government records ([Romer and Romer \(2010\)](#)). Relying on natural experiments provides an alternative way of establishing exogeneity: for example, a war initiated by another country may create a natural experiment that causes increased government spending not motivated by current economic conditions in the home country. Although the absence of reverse causality might be easier to argue, the approach still faces some important challenges. First, one needs a critical number of these events over time or geographical variation in order to carry out an empirical analysis. A further hurdle lies in controlling for effects of the “natural experiment” on economic growth that do not play out via government spending. For example, a war might affect patriotism in the home country, potentially increasing the demand for home-produced goods, or a war taking place in the home country likely affects the capital stock. We review two lines of this literature: the first one relies on exogenous variation in military spending, and the second one estimates local fiscal multipliers, relying on different natural experiments.

3.1 Permanent Income Hypothesis Studies and the Fiscal Multiplier

Before analyzing the use of natural experiments in establishing exogeneity of fiscal spending, we want to point out the intimate link between the literature on the Permanent Income Hypothesis and the question of the size of the fiscal multiplier. Some of the

Permanent Income Hypothesis papers involving natural experiments mentioned above lend themselves naturally to answering questions about the effectiveness of fiscal policy. Are tax rebates an effective means to stimulate the economy? The answer depends on whether households save or spend the tax rebates that they receive. However, answering this question with the studies above has four important caveats: the Permanent Income Hypothesis would predict that households adjust their consumption at *announcement* of the stimulus, whereas the papers cited above analyze consumption reaction at the *implementation* of the stimulus. In that sense, using them to test the Permanent Income Hypothesis and to at the same time analyze the response of household consumption to an economic stimulus is inconsistent. On the other hand, one can consider the estimates found in these papers as a lower bound of spending, because households could have adjusted their consumption already partly at the announcement or during the time between announcement and implementation. Agarwal and Qian (2014a), which is the only paper that explicitly analyzes consumption reactions to a temporary income increase both at announcement and at implementation, find a significant consumption response already at announcement, which carries over to the time period after receipt of the payment. Second, although the receipt of a rebate check can be considered exogenous for an individual household, for the economy as a whole, it is certainly not exogenous. For example, the 2008 Economic Stimulus Payments were explicitly designed to stimulate the economy. Third, the papers analyze only a partial equilibrium response of households, not taking into account any general equilibrium effects. Fourth, most of the cited studies analyze expenditure on non-durables, whereas for the fiscal multiplier, total spending matters.

That said, the studies that analyze fiscal policy measures find evidence for a large spending response by households. Parker et al. (2013) find that households spent between 50 percent and 90 percent of the 2008 Economic Stimulus Payment on durable and non-durable goods in the quarter following receipt, indicating that the majority of the payments were consumed, not saved. Johnson et al. (2006) find that non-durable consumption increased by 20 percent to 40 percent of the payments in the quarter following receipt of the 2001 tax rebates, and that around two thirds of the rebates were spent in total on nondurable consumption, considering the cumulative effect over the six months period following receipt.²⁵ Misra and Surico (2014) stress the heterogeneity of the consumption response to both the 2001 and 2008 episodes. They show that the aggregate consumption response is smaller once this heterogeneity is taken into account than the

²⁵They do not analyze expenditure on durable goods.

estimates imposing homogeneity would suggest. The total disbursements of \$38 billion in 2001 and \$96 billion in 2008 lead to an estimated aggregate consumption response of \$16 (\$26) billion in non-durable consumption in 2001 based on the heterogeneous (homogeneous) estimates, and to an increase of \$15 (\$56) billion in total consumption in 2008.²⁶ [Sahm et al. \(2010\)](#) use survey data on spending intentions. Their estimated spending responses are smaller than the ones estimated from actual expenditure data, indicating that around one third of the 2008 Economic Stimulus Payments were spent.²⁷ Analyzing an older episode, [Hausman \(2015\)](#) finds that within one year, World War I veterans spent between 60 percent and 75 percent of bonuses that they received in 1936, mostly on cars and housing.

3.2 Military News Shocks as Natural Experiments

One way to address the potential endogeneity of government spending is to use military events as natural experiments. The work of [Barro \(1981\)](#) and [Hall \(1986\)](#) recognized the usefulness of military events in this regard early on. Geopolitical events leading to a large buildup of military expenditure are often plausibly exogenous, because they arise due to actions of some other nations. Thus, they can potentially be used as natural experiments to isolate exogenous changes in government spending. The first paper systematically following this approach is the study by [Ramey and Shapiro \(1998\)](#). Based on newspaper articles, they identify three major military news shocks in the post-World-War II area: the Korean War news shock in the third quarter of 1950, the Vietnam War news shock in the first quarter of 1965, and the Carter-Reagan buildup after the Soviet invasion of Afghanistan in the first quarter of 1980. [Ramey \(2011\)](#) adds to this list the shock of September 11, 2001. Because the identification is based on newspaper articles, these studies also fall under the “narrative approach,” but they try to identify the dates of military shocks that can be used as natural experiments. Note that timing the news shock is not trivial, especially during the Vietnam War. [Ramey and Shapiro \(1998\)](#) document that newspaper articles only started arguing about a military buildup after the February 1965 attacks on the US Army barracks, long after the military coup of November

²⁶[Kaplan and Violante \(2014b\)](#) point out potential reasons for the generally lower impact of the 2008 reimbursements, among others the larger size of the individual household transfers, and the phasing out at the lowest income levels.

²⁷Using the same methodology and data, [Sahm et al. \(2012\)](#) find that the form of payment matters, and reducing tax withholdings leads to a smaller consumption response than explicitly distributing a rebate. [Misra and Surico \(2014\)](#) provide some reconciliations of the different estimates of consumption responses based on expenditure data and survey responses.

1, 1963, in Vietnam. Military actions of foreign entities caused all four events, such that the argument that they were exogenous to current economic conditions in the United States is very plausible.

However, using these events as natural experiments to analyze the size of the fiscal multiplier still poses some challenges. Specifically, these news might affect other relevant variables that influence GDP, rather than only government spending. This argument especially holds for World War II, in which rigid price controls were introduced and patriotism was strong, both of which might have had a direct effect on labor supply. For this reason, [Ramey and Shapiro \(1998\)](#) exclude World War II from the analysis. A thorough discussion of potential confounding factors in the other episodes is, however, somewhat missing from the literature. For example, the terrorist attacks of 9/11/2001 likely affected uncertainty about the future path of the economy, and uncertainty shocks can matter for economic performance (see, e.g. [Bloom \(2009\)](#)).

The early study by [Ramey and Shapiro \(1998\)](#) runs a regression of GDP growth on these quarterly military event dummies with up to eight lags, whereas [Ramey \(2011\)](#) uses these dummies in a VAR approach. She finds that the military event dummies significantly precede increases in military spending, and concludes that traditional VAR approaches might fail to identify anticipation effects. This failure can potentially reconcile the different outcomes that the traditional VAR approach and the “narrative approach” relying on military buildups have found in the effect of government spending on private consumption and real wages: traditional VAR approaches typically find a positive effect of government spending shocks on consumption and the real wage, whereas the papers relying on military news shocks find a negative effect on these two variables. [Edelberg et al. \(1999\)](#) provide two robustness exercises for [Ramey and Shapiro \(1998\)](#): first, they show that results are robust to small disturbances in the event dates. Second, and more importantly, they run placebo exercises in the spirit of the placebo exercises discussed in [Section 2.2.1](#) and find that using arbitrary event dates leads on average to response functions outside of the confidence bands of the true responses.

[Ramey \(2011\)](#) goes one step further than relying simply on dummy variables, and constructs a defense news variable that measures the change in expected net present value of future military spending at the quarterly frequency based on newspaper accounts. This variable is a strong predictor of government spending as long as World War II is included. The implied fiscal multiplier relying on a VAR estimation lies between 1.1 and 1.2, but falls

to between 0.6 and 0.8 if World War II is excluded.²⁸ In line with this evidence, [Hall \(2009\)](#) stresses that the identifying variation in these studies comes from large wars, especially World War II and the Korean War. Focusing on differential effects by spending categories, [Auerbach and Gorodnichenko \(2012\)](#) find that military spending is the spending variable associated with the largest multiplier.

3.3 Local Fiscal Multipliers

The literature on fiscal multipliers has recently started using natural experiments to establish the exogeneity of the fiscal instrument by relying on regional data and estimating local fiscal multipliers. These multipliers are local in the sense that they analyze the effects of changes in local spending financed by the federal administrative level, which therefore constitute windfall payments from the point of view of the localities and are not associated with an increase in taxation or local debt.²⁹ Importantly, this approach ignores general equilibrium effects at the national level. [Nakamura and Steinsson \(2014\)](#) provide a theoretical discussion of how the local fiscal multiplier estimate can be tied to a standard aggregate fiscal multiplier; we come back to this shortly at the end of this section.

A typical regression run in the local fiscal multiplier literature is the following variant of (9):

$$\Delta Y_{i,t+1} = \alpha + \beta \Delta \hat{F}_{i,t+1} + \gamma' \Delta X_{i,t+1} + \delta_i + \eta_t + \epsilon_{i,t+1}, \quad (10)$$

where subscript i stands for the local entity, δ_i are regional fixed effects, and η_t are year fixed effects.³⁰

An advantage of the local fiscal multiplier regression (10) over specification (9) is that it allows the inclusion of regional and year fixed effects. The regional fixed effects capture any time-invariant regional characteristics that could lead to systematically lower or higher growth in a respective region (e.g. in urban vs. rural regions).³¹ More important in

²⁸In [Owyang et al. \(2013\)](#), this defense news variable is constructed for the United States from 1890 to 2010, and for Canada from 1921 to 2011. The authors then also analyze whether the government spending multiplier is larger in recessions than in booms, and find evidence in favor for Canada, but not for the United States. [Barro and Redlick \(2011\)](#) also use this variable on an annual frequency to estimate multipliers, and find somewhat smaller multipliers than in [Ramey \(2011\)](#).

²⁹An exception to this rule is the paper by [Clemens and Miran \(2012\)](#).

³⁰Note that [Corbi et al. \(2014\)](#) and [Shoag \(2013\)](#) instead regress output growth on the level of the fiscal instrument, rather than its change.

³¹The papers by [Serrato and Wingender \(2014\)](#) and [Chodorow-Reich et al. \(2012\)](#) are exceptions in this regard by not including regional fixed effects. The former instead includes state-decade fixed effects, with the local level being the county level. The latter one does not include a time dimension, thereby

the context of estimating the fiscal multiplier might, however, be the inclusion of year fixed effects. These effects allow one to control for any national fiscal and more importantly monetary policies that happen concurrently with the local fiscal policy. Monetary policy is often correlated with fiscal policy, and disentangling the effects of both is therefore a major challenge for any macro estimation of fiscal multipliers. However, because monetary policy is conducted exclusively on the national level, it can easily be controlled for by including year fixed effects in an estimation of local fiscal multipliers. Given that local fiscal multiplier studies all analyze multiple local subentities of a country, they lend themselves naturally to analyzing potential heterogeneous effects of fiscal multipliers depending on local characteristics such as business cycle conditions, openness, financial development, and so on.

Instrumental Variables To address potential endogeneity of the fiscal instrument, regression (10) is estimated via instrumental variables, where $\Delta \hat{F}_{i,t+1}$ is the predicted change in fiscal spending based on a first-stage regression involving the instrument I :

$$\Delta F_{i,t+1} = \kappa + \theta I_{i,t} + \zeta' \Delta X_{i,t+1} + \varepsilon_{i,t+1}. \quad (11)$$

The challenge of the estimation is to find a valid instrument for the fiscal measure. The exclusion restriction for the instrument is that it affects output growth only through its effect on the fiscal measure. Here is where natural experiments step in. Consider, for example, the paper by [Serrato and Wingender \(2014\)](#). Their instrument relies on the fact that federal spending at the local level is tied to the size of the local population. The estimates of the local population size come from different sources in different years: the census carried out every 10 years provides direct counts of the local population, whereas in the years in between two censuses, population estimates are updated based on vital statistics and estimated migration flows. As a result, substantial fluctuations occur in measured population in the year before a decennial census and the census year, which are called census “error of closure.” The authors use this census error of closure to instrument the change in federal spending on the local level in the affected years.

For the census error of closure to be a valid instrument, it has to predict fiscal spending, and the exclusion restriction has to hold. The first of these two conditions can relatively easily be established by showing the first-stage regression results and running an F-test for the joint significance of the instruments. Concerning the second condition, the authors

preventing the use of local fixed effects.

show theoretically that under classical measurement error in both census counts and administrative estimates, or under the weaker condition that both estimates are biased in the same order of magnitude, the exclusion restriction holds. Moreover, they explain in detail how the two estimates arise and what the literature concludes on their accuracy and potential biases. For example, one could imagine that the population estimates between censuses systematically underestimate population growth in fast-growing counties, such that the error of closure is always more positive in counties that experience an economic expansion. Any persistence in growth would then result in a direct effect of the error of closure on growth, violating the exclusion restriction. Controlling for past growth helps address this concern but does not rule it out completely. To provide further evidence that the exclusion restriction holds, the authors show that the census error of closure shows only minimal geographical correlation at the county level and no time-series correlation. Most importantly, it is not positively correlated with growth in the years before the error of closure should actually affect federal fiscal spending on the county level.³² Although in the end these exercises remain suggestive, addressing potential concerns about the exclusion restriction in further evidence of this kind is good practice.

This literature uses two other interesting natural experiments. The first one, used in the paper by [Acconcia et al. \(2014\)](#), is an Italian law specifying the dismissal of elected local officials and their replacement with three external commissioners for 18 months upon evidence of Mafia infiltration in city councils. This replacement leads on average to sharp reductions in spending on public work at the provincial level, the reason being that this sector is typically a lucrative source of business for the Mafia. The authors show that growth rates prior to dismissal are not significantly different in treated and control provinces. The second natural experiment in the paper by [Cohen et al. \(2011\)](#) consists of changes in congressional committee chairmanships, which influence government spending in the state of the new chairman.³³ Because chairmanship in a committee is largely determined by seniority, and because chair turnover results from election defeat or resignation of the incumbent, it is driven by political circumstances in other than the home state of the incoming chairman and can thus be seen as exogenous.

³²The error of closure should only affect fiscal spending two years after the census is run, given that publishing the results takes two years. In fact, a significant negative correlation exists between the error of closure and employment and income growth in previous years. This finding might raise the worry that spending rises in past recession areas, and that mean reversion might lead to future growth in these areas. The authors argue that controls for past growth in the second-stage regression take care of these concerns.

³³[Feyrer and Sacerdote \(2012\)](#) take a similar approach, relying on average seniority of House members, when analyzing the effectiveness of the 2009 ARRA.

There exist other papers in this literature that rely on the same approach of finding a valid instrument for the fiscal measure in equation (10), but in which this instrument is less clearly a natural experiment. There is an obvious “grey zone” of what can be considered a natural experiment. In the spirit of our definition that a natural experiment is an historical event that provides exogenous variation to give a plausible identifying assumption, the Census error of closure, the law specifying replacement of local officials upon evidence of Mafia infiltration, and the chairmanship in congressional committees, are such historic episodes. A variety of papers in this literature use instruments to identify exogenous variation in government spending, where the instruments do not rely on historical episodes and are themselves more directly linked to fiscal policies. Examples of these papers are [Nakamura and Steinsson \(2014\)](#), who exploit different state-level sensitivities to national military spending,³⁴ [Chodorow-Reich et al. \(2012\)](#), who use pre-crisis state-level Medicaid spending to extract the exogenous component of increases in federal match components of state Medicaid expenditure during the 2009 American Recovery and Reinvestment Act (ARRA), and [Wilson \(2012\)](#), who uses exogenous formulary allocation factors such as federal highway miles in a state or a state’s youth share to instrument government spending under the 2009 ARRA. Similarly, [Clemens and Miran \(2012\)](#) use fiscal institutions on the state level, specifically how strict balanced budget rules are, to identify exogenous variation in government spending. [Shoag \(2013\)](#) and [Shoag \(2015\)](#) exploit variations in returns to state pension plans.³⁵ [Kraay \(2012\)](#) and [Kraay \(2014\)](#) focus on World Bank lending, in which project financing is spread out over several years after the lending decisions. In all of these cases, the instrument is quite closely related to the research question, which makes it harder to argue that the exclusion restriction holds. Therefore, all of these papers spend considerable effort on providing additional evidence for the validity of the exclusion restriction. A common strategy to establish exogeneity of the instrument is to include further controls and exploit specifics in the timing to argue that the identifying assumption holds; see, for example, [Clemens and Miran \(2012\)](#)

³⁴[Fishback and Cullen \(2013\)](#) analyze the effect of state-level military spending during the Second World War, but do not use instruments for state-level military spending, but rather rely on narratives to establish exogeneity.

³⁵Both papers use the same instrument. However, the analysis in [Shoag \(2013\)](#) focuses on the Great Recession years 2008 and 2009, whereas [Shoag \(2015\)](#) exploits information from 1987 to 2008. [Brückner and Tuladhar \(2014\)](#) also analyze a local fiscal multiplier using data from Japanese Prefectures. Whereas their study shares the use of year and region fixed effects with the studies cited above, it addresses the issue of endogeneity of local government expenditures by using a system GMM approach. The paper by [Fishback and Kachanovskaya \(2015\)](#) analyzes federal spending on the state level during the Great Depression, relying on instruments similar in spirit to [Nakamura and Steinsson \(2014\)](#), [Chodorow-Reich et al. \(2012\)](#), and [Wilson \(2012\)](#)

and Kraay (2012).³⁶ Wilson (2012) controls for a variety of variables that are potentially correlated with post-2009 growth and also 2009 ARRA spending, for example, pre-2009 employment growth.

Regression Discontinuity There are only few papers in the local fiscal multiplier literature using a regression discontinuity approach. Since the exploited policy rules that generate the discontinuities are close to the research question at hand, these papers are closer in spirit to the second group of papers cited above than to the first group, which relies more obviously on natural experiments. Corbi et al. (2014) exploit the fact that, as in the United States, Brazilian federal transfers to municipal governments rely on the population at the local level. In contrast to the United States, a specific step function exists that specifies the total transfer amount for certain population classes. Thus, sharp discontinuities arise in the transfers per capita around the cut-off values in this step function, whereas all other variables should change smoothly around the cut-off. This is the identifying assumption for their regression discontinuity approach. Another advantage of their experiment is that several cut-off values exist (rather than, e.g., only one cut-off value as in the paper by Trezzi and Porcelli (2014)), which gives the test high statistical power. Sixty percent of the municipalities in the sample switch the population class at least once in the sample period. Because the data show that adherence to the cut-off is not implemented 100 percent, and, in fact, some judiciary disputes surround them, the authors confront a fuzzy regression discontinuity design and use the theoretically predicted transfers based on the actual population count as an instrument for the actual transfers. They employ different bandwidths around the cut-off values, and also report results from regressions with a rectangular kernel. In a similar spirit, Becker et al. (2010) and Becker et al. (2013) use a regression discontinuity design to analyze the effect of EU grants on local development, relying on a discontinuity in regional GDP per capita eligibility for the grants.

Estimates of the Local Fiscal Multiplier The estimates of the local fiscal multipliers in the studies described above show surprising consistency and range between 1.5 and 2, despite the different identifying restrictions based on natural experiments or other institutional details, different econometric approaches, different nations, and different

³⁶Kraay (2012) in addition uses predicted project-level disbursements, given the economic sector and the geographic region, rather than actual ones to address concerns about endogeneity.

fiscal spending measures.³⁷ Nakamura and Steinsson (2014) show that an open economy New Keynesian model with sticky prices and complementarity between consumption and labor is best able to replicate these large local fiscal multipliers, which they call “open economy relative multiplier”. They also show that the model implied “closed economy aggregate multiplier”, which is commonly considered in the literature on fiscal multipliers, depends in its size very much on the accompanying monetary policy.

Another common feature of these studies is that the IV estimates are 5 to 15 times larger than the simple OLS estimates of equation (10).³⁸ This indicates that OLS estimates might be systematically downward biased, for example, because of automatic stabilizers (which increase spending in downturns of the economy), general endogeneity of government spending (increased discretionary spending to stimulate the economy in downturns), and interaction with monetary policy (fiscal spending might step in specifically when monetary policy does not work due to a binding zero lower bound).

4 Identification: Causal Factors in Economic Growth

4.1 The Fundamental Causes of Growth

In the previous two sections, we have explored a number of natural experiments that inform us about the validity of established macroeconomic models and the direction and magnitude of causal relationships that operate within them. In this section, we ask what natural experiments can teach us about the kind of models that we *should* be writing. In other words, what are the important determinants of economic outcomes that are outside of our standard models?

By far the most influential application of natural experiments in this respect is the search for the “fundamental” reasons of why some countries are rich while others are poor. Corresponding to the breadth of this question, the empirical work in this area is very diverse and much less focused than the work we covered in the previous sections. For this reason, we find it useful to first lay out a rough framework to fix ideas and organize the different strands of this literature into a coherent narrative.

To distinguish the fundamental causes from proximate causes of economic growth,

³⁷These multipliers to windfall income are larger than the multiplier estimated by Clemens and Miran (2012), who take local tax or debt adjustments into account.

³⁸For example, in Serrato and Wingender (2014), the 2SLS estimate is 15 times larger than the OLS estimate, 7 times larger in Acconcia et al. (2014), and 5 (based on regional data) to 15 (based on state data) times larger in Nakamura and Steinsson (2014).

consider a typical production function that relates output Y_t to the input of physical capital K_t , human capital H_t , the number of workers L_t , and the level of labor-augmenting technology A_t :

$$Y_t = K_t^\alpha H_t^\beta (A_t L_t)^{1-\alpha-\beta}, \quad (12)$$

where $0 < \alpha < 1$, $0 < \beta < 1$, and $\alpha + \beta < 1$. By dividing both sides of this expression by L , we can see that output per capita is a function of technology and the intensity of physical and human capital per worker. Equation (12) thus describes a causal relationship: having better technology and more capital and education per worker makes a country richer.³⁹

Although this statement is undoubtedly true, it is also not very helpful for understanding the large and persistent differences in income per capita across countries, without understanding the process by which countries accumulate capital and technology. This question is the object of a large theoretical literature on economic growth.

Consider, for example, the canonical growth model based on Solow (1956) and Swan (1956). In this model, A_t grows at the exogenous rate g and at every point in time, and households invest a fixed proportion of their income into physical and human capital according to

$$\dot{K}_t = s_k Y_t - \delta_k K_t$$

and

$$\dot{H}_t = s_h Y_t - \delta_h H_t,$$

where the savings rates s and depreciation rates δ are between zero and one, and $s_k + s_h < 1$.

Using lowercase variables to denote the level of output, physical, and human capital in terms of effective labor,

$$x_t \equiv X_t / (A_t L_t), X = Y, K, H,$$

we can then show that at the balanced growth path, consumption, output, physical, and human capital per effective unit of labor are all constant. In particular, output per effective labor is

$$y^* = \left(\frac{s_k}{g + \delta_k} \right)^{\frac{\alpha}{1-\alpha-\beta}} \left(\frac{s_h}{g + \delta_h} \right)^{\frac{\beta}{1-\alpha-\beta}}.$$

It follows directly that output per capita (as well as physical and human capital per

³⁹Here we are assuming that all countries share the same production function (12) but do not otherwise interact with each other.

capita) grows at the fixed rate g .

The Solow-Swan model again describes a causal relationship: having better technology and saving more for investment in human and physical capital makes a country richer. Although this relationship gives us a better idea of *how* growth happens (by saving and accumulating capital), it does arguably little for our understanding of *why* people in in a poor country like Zimbabwe use worse technology and/or invest less in physical capital and education than people in a rich country like the United States.

In the half century since the publication of the [Solow \(1956\)](#) and [Swan \(1956\)](#) papers, growth theory has made tremendous advances in understanding the mechanics of growth. Aside from endogenizing the savings rates (s), one of the major advances of the literature has been to explain the dynamics of the technology process, such that A_t becomes effectively an endogenous accumulated factor itself. With one major exception, which we will discuss in detail below, these advances, however, do not change the fundamental nature of the problem: physical capital, education, and technological progress are proximate rather than fundamental causes of growth. Understanding their dynamics helps us understand the mechanics of growth (how it happens) but conceptionally does not tell us why a poor country like Zimbabwe is fundamentally different from a rich country like the United States.

For the purposes of our following discussion, we will define the fundamental causes of economic growth as *the political, institutional, and social reasons that are preventing many countries from investing enough in technology and productive factors*. Formally, we may think of a typical growth model as producing a mapping from a vector of parameters ϕ_i governing the production and accumulation of technology and productive factors in a country i to its level of output per capita $\frac{Y_{i,t}^*}{L_{i,t}}$ on the balanced growth path:

$$F(\phi_i) \rightarrow \frac{Y_{i,t}^*}{L_{i,t}}. \quad (13)$$

Whereas we think of the proximate causes of growth as operating within the function F , the fundamental causes are the political, institutional, and social factors that generate cross-country variation in the parameters ϕ_i .

In the remainder of this chapter, we consider the evidence that natural experiments have uncovered for three such fundamental causes:

1. institutions,
2. social structure, and

3. culture.

For the purposes of our following discussion, we define as “institutions” the broad set of rules, regulations, laws, and policies that affect economic incentives and thus the incentives to invest in technology, physical capital, and human capital (Acemoglu, 2009). By social structure we mean “patterned relations,” that is, the network of friendships and family ties between large groups of individuals that affects the diffusion of information and the ability of individuals to enforce contracts, as well as the system of socioeconomic stratification (e.g., the class structure) of a society. As “culture” or “civic capital,” we define those persistent and shared beliefs and values that help a group overcome the free-rider problem in the pursuit of socially valuable activities (Guiso, Sapienza, and Zingales, 2012). Note that the literature often refers to the latter two categories jointly as “social capital” (Putnam, 2000). We prefer to separate them here because both of these elements have been empirically shown to have independent effects on growth, but otherwise have no preference for one definition over another.

Taken together, these three causes determine the social environment in which economic activity takes place. Because conducting controlled experiments on this social environment at the scale of countries or even regions is generally impossible, natural experiments are arguably our best bet for making causal inference at this scale. In each case, we will first discuss a selection of natural experiments that provide evidence of the causal relationship between GDP per capita and institutions, social structure, and culture, respectively. Then, whenever available, we will discuss some of the evidence on the dynamics of each of these forces and avenues for future research. Because each of these literatures is vast, we cannot give a comprehensive overview of each of them (Guiso et al. (2012), Algan and Cahuc (2013), Alesina and Giuliano (2015), Nunn (2013), and Chaney (2014) provide excellent surveys). Instead, we focus on giving a broad overview of the types of measurement and identification issues that arise in these applications, and provide specific examples.

To our list of three fundamental causes, we add a set of natural experiments that speak to the question of whether multiplicity might exist in the mapping (13). In other words, the mechanics of economic growth may be such that two countries with identical ϕ_i may end up on different balanced growth paths by pure coincidence or luck. This scenario is the major exception mentioned above: models that generate multiple equilibria may be able to explain why Zimbabwe is poor and the United States is rich, without taking recourse to any fundamental causes of growth:

4. luck and multiple equilibria.

Here again we select papers to give a broad overview of the measurement and identification issues.

One striking feature that the studies in all four categories have in common is that they tend to make relatively little use of the structure provided by growth models. That is, most approaches relate variation in ϕ_i directly to $\frac{Y_{i,t}^*}{L_{i,t}}$, without making use of the structure of F . In this sense, the literatures on the proximate causes and fundamental causes of growth have developed relatively independently. We will comment on this issue further below.

4.2 Institutions and Political Economy

Of the three fundamental causes of economic growth, “institutions”—the rules, regulations, laws, and policies that affect the incentives to invest in accumulated factors of production—have received by far the most attention from empiricists. The reason for this attention is that institutions, and in particular the protection of property rights, fit neatly into standard ways of thinking about incentives: investors will be reluctant to invest in capital and technology if the returns from these investments are likely to disappear into the pockets of corrupt bureaucrats. Clearly, bad institutions must thus be bad for growth. A correspondingly long history of thought in economics links institutions to economic development (North and Thomas, 1973; Jones, 2003; North, 1981; Mauro, 1995; Hall and Jones, 1999; La Porta, Lopez-de-Silanes, Shleifer, and Vishny, 1998).

However, despite this long intellectual history, which way the causality runs is far from clear: for example, we may suspect that richer countries can also afford a better judicial system, and thus better protect property rights. Similarly, researchers disagree about what makes governments *want to* protect property rights. In practice, the set of institutions a society adopts depends in complex ways on the society’s level of economic development, culture, social structure, the education of the population, and its historical experience. Identifying a causal effect of institutions on economic growth thus requires an exogenous source of variation in the type of institutions that different societies adopt.

Consider, for example, the following structural equation of interest:

$$\log(Y_i) = \alpha + \beta R_i + X_i' \gamma + \epsilon_i, \tag{14}$$

where Y_i is income per capita in country i , R_i is a measure for the protection of property

rights, and X_i is a vector of controls. The fundamental problem in identifying β is that R_i and Y_i are likely to be jointly determined such that $cov(R_i, \epsilon_i) \neq 0$, resulting in biased OLS estimates of (14). Unbiased estimates of β thus require exogenous variation in R_i .

In this section, we first discuss a set of natural experiments that attempt to identify such exogenous variation, and establish a causal link between institutions and growth. We then turn to the mechanism that links institutions to growth and to the dynamics of institutions.

4.2.1 The Effect of Institutions on Growth

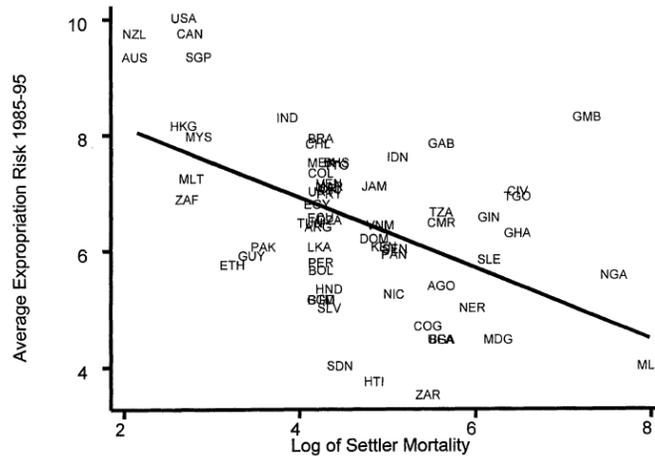
In an influential study, [Acemoglu, Johnson, and Robinson \(2001\)](#), consider the natural experiment surrounding the colonization of virtually all of Africa, North and South America, Australia, and large parts of Asia by Europeans. Today, the countries that emerged from these former colonies differ widely in their level of economic development and in the functioning of their institutions – and these differences appear highly persistent over time. For example, the United States has been richer than the Congo for more than a century, and there is little evidence that this will change in the near future. Acemoglu, Johnson, and Robinson (AJR) argue that the mortality rates Europeans settling in these different colonies faced explain part of this heterogeneity: Europeans who made the trip to the Congo could expect to be plagued by malaria and yellow fever and consequently had much lower life expectancies than settlers heading for the present-day United States. This health cost that Europeans faced when considering whether to settle in different parts of the world determined the form of colonization that each country experienced.

AJR’s idea is that in countries with a disease environment that was hazardous to Europeans (e.g., the Congo), the early colonizers set up institutions that would maximize the extraction of natural resources, requiring as small a European presence as possible. These institutions did not introduce much protection of property rights and did not involve checks and balances against the power of the ruling elites that were dominated by Europeans. By contrast, colonies that were attractive destinations for European settlers (e.g., Australia and the United States) imported European institutions that fostered safe property rights and checks and balances in government. The resulting historical differences in institutions across countries then tended to persist to the present day, resulting in persistent differences in economic performance that last to the present day.

Figure 2 shows the relationship between settler mortality and a contemporary measure for average expropriation risk for the 64 countries included in AJR’s sample. The authors

argue that this negative and highly significant relationship delivers quasi-random variation in R_i because the mortality rates of European settlers in the 17th to 19th centuries (or any omitted variables correlated with them) are unlikely to have a direct effect on GDP per capita today, other than their effect through institutions.

Figure 2: Relationship between settler mortality and expropriation risk in [Acemoglu et al. \(2001\)](#)



When instrumenting R_i with the mortality rates of soldiers, bishops, and sailors stationed in country i between 1600 and 1800 CE, AJR's estimates of the parameter β are positive and statistically highly significant. They imply that an improvement in Nigeria's level of protection of property rights to match that of Chile would, in the long-run, increase Nigeria's GDP per capita by a factor of 7. Heterogeneity in institutions may thus account for a large part of the differences that exist today in wealth between countries.

In an additional application of the same natural experiment, [Acemoglu, Johnson, and Robinson \(2002\)](#) argue that heterogeneity in institutions can also explain the large changes in the relative prosperity that have occurred since 1500 CE. Before the onset of European colonization, civilizations in Meso-america, the Andes, India, and Southeast Asia appear to have been richer than those in North America, Australia, or New Zealand. Today, the relationship is reversed: a simple regression of GDP per capita in 1995 on urbanization in 1500 (as a proxy for historical GDP per capita) yields a negative and highly statistically significant coefficient. AJR show that differences in the institutions imposed by European colonizers can again explain this fact. When adding measures of institutions to the regression and instrumenting them with settler mortality, the negative

correlation between present-day GDP per capita and historical urbanization disappears. Because settler mortality and population density in 1500 are positively correlated, European colonizers tended to introduce or perpetuate extractive institutions in places where they found a high density of local populations, while creating good institutions in places where they settled in large numbers.⁴⁰

The main identifying assumption for a causal interpretation of the results in both papers is that, conditional on the controls included in the vector X , the mortality rates of European settlers in the 17th to 19th centuries (or any omitted variables correlated with them) have no effect on GDP per capita today, other than their effect through institutions.

This assumption raises two main concerns. First, omitted variables that are correlated with historical mortality rates could be correlated with present-day GDP per capita. The most obvious of these variables is the current disease environment. For example, places that were hard to settle historically may be difficult to live in today, resulting in lower economic growth. AJR argue that this is not the case. The historical disease environment had large negative effects on Europeans, but much smaller effects on the indigenous populations. For example, local troops serving the British army in Bengal had mortality rates that were comparable to those of British troops stationed in Britain, whereas British troops in Bengal had mortality rates that were 7 to 10 times higher. In addition, AJR's estimates for β change little when controlling for elements of the current disease environment, suggesting that the effect of the historical disease environment indeed transmits itself through differential European settlement at the time. They also show that their results are robust to controlling for other correlates of GDP per capita, such as the identity of the colonizer, the origin of the legal system, and various measures of climate conditions.

The second main concern is much harder to address: the effect of European settlement may transmit itself through a range of different mechanisms that may be correlated with the protection of property rights. In particular, Europeans may also have imported their own culture and high levels of civic capital or they may have brought valuable social ties to their countries of origin. Thus, AJR are unable to distinguish the effect of institutions from the effect of other persistent variables that may be correlated with both European settlements and institutions, and in particular from the two other fundamental causes of

⁴⁰For a more in-depth discussion of the merits and weaknesses of AJR (2001)'s analysis, see Easterly and Levine (2003, 2012), Glaeser, La Porta, Lopez-De-Silanes, and Shleifer (2004), Olsson (2004), Rodrik, Subramanian, and Trebbi (2004), Albouy (2012), Acemoglu, Johnson, and Robinson (2012), and Acemoglu, Gallego, and Robinson (2014).

economic growth discussed in the next two subsections. More generally, any instrumental variables approach that relies on cross-sectional variation will run into this problem.

Michalopoulos and Papaioannou (2014) attempt to make progress on distinguishing the effects of institutions and culture using data on light intensity in combination with a regression discontinuity approach. They argue that the borders between many African countries were drawn in the mid to late 19th century by Europeans who were largely uninformed about local conditions. As a result, these borders partition more than 200 historic homelands of ethnicities between two different modern-day countries in a quasi-random way, subjecting identical cultures residing in geographically homogeneous territories to different country-level institutions. Their main specification takes the following form:

$$y_{p,e,i} = \alpha + \beta IQL_i^{HIGH} + f(BD_{p,e,i}) + \gamma PD_{p,e,i} + X'_{p,e,i} \Phi + \eta_e + \epsilon_{p,e,i}, \quad (15)$$

where $y_{p,e,i}$ is a dummy variable that is 1 if pixel p (an area corresponding to about 12×12 km) in the historic homeland of ethnic group e in country i is lit — a simple measure of whether residents of that piece of land can afford nighttime lighting. IQL^{HIGH} is an indicator that is 1 if the pixel is in the part of the partitioned homeland that is located in the country with the relatively higher institutional quality, $f(BD_{p,e,i})$ is a polynomial of the shortest distance of the pixel centroid to the country border, PD is population density, X contains additional controls, and η_e is an ethnic homeland fixed effect. Under the identifying assumption that at the country border, institutions change discretely, while all other relevant influences on $y_{p,e,i}$ (including culture and geography) change continuously, β measures the effect of relatively better institutions at the country border.

In contrast to the results of AJR (2001)'s cross-country analysis, the authors find no effect of institutions at the country border: across different variations, they cannot distinguish β from zero, suggesting that, at the border, exogenous variation in national institutions is not associated systematically with higher wealth (more light). These results may suggest that, at least for Africa, omitted factors such as culture (civic capital) or social structure might explain the strong association between institutions and GDP per capita. However, more consistent with AJR (2001)'s results, Michalopoulos and Papaioannou find a positive and significant effect of better institutions for split groups that are located close to national capitals. The absence of an effect at the border of the average African country could thus simply be explained by the fact that the influence of many African countries'

governments is weak in remote regions.⁴¹

In a closely related paper, [Pinkovskiy \(2013\)](#) also uses a regression discontinuity approach to test for discontinuities in the amount of light per capita at country borders, but applies his analysis to the entire world, rather than only to Africa. Because most borders in the world were not drawn randomly but might be determined by the location of ethnic homelands, culture, or other variables, he restricts his estimates to within 50km of the border. Within this narrow bandwidth, one may plausibly argue that the *exact* location of the border is quasi-random, even though borders outside of Africa were not drawn by largely uninformed imperial powers. In contrast to [Michalopoulos and Papaioannou \(2014\)](#), Pinkovskiy finds a large and statistically significant discontinuity at national borders, although it is not significant when restricting the sample to Africa. In addition, this effect of the country border becomes statistically insignificant once Pinkovskiy controls for differences in the rule of law.

4.2.2 The Effect of Institutions on Business Cycles and Conflict

Given this evidence of a causal link between institutions and growth, an obvious question is *how* this effect transmits itself in practice. One interesting, and often overlooked, piece of evidence comes from yet another application of AJR (2001)'s natural experiment. [Acemoglu, Johnson, Robinson, and Thaicharoen \(2003\)](#) show that countries with worse institutions (again instrumented with settler mortality) also have more volatile business cycles and are more prone to episodes of economic crises. Once the authors control for this effect of institutions, the standard macroeconomic policies that are often blamed for macroeconomic instability (e.g., high government spending and high inflation rates) are no longer systematically associated with macroeconomic volatility.

The authors draw two main conclusions from this finding. First, bad macroeconomic policies are often a symptom of underlying institutional problems. Second, they are often not the primary mediating channel through which institutions affect macroeconomic stability. These results suggest that a more useful line of thought might be that policies such as excessive government spending and high inflation rates are just two of a large set of tools that politically powerful groups use to extract rents from the economy. Any policy (perhaps imposed by the IMF or another international organization) that shuts down one of these distortions but does not resolve the underlying institutional problems, may then

⁴¹There is also a technical debate on the bleeding of nighttime light across borders that may attenuate any estimated border-effect, see the discussion in [Pinkovskiy \(2013\)](#).

simply result in the use of a different tool (another macroeconomic or microeconomic distortion) that achieves the same aim. Macroeconomic instability and low growth may then be thought of as collateral damage from this rent-extraction process.⁴²

In a second application of their natural experiment surrounding the quasi-random drawing of African boundaries by relatively un-informed Europeans, [Michalopoulos and Papaioannou \(2011\)](#) show that ill-designed institutions may also slow economic development by prompting long-term conflict. Using data on the pre-colonial locations of 834 ethnic groups, they show that the drawing of African boundaries indeed appears to have quasi-randomly partitioned 229 of these groups between two adjacent states: partitioned and non-partitioned ethnicities do not appear to differ systematically in their pre-colonial characteristics, except that partitioned groups tended to cover larger land areas and to have historical homelands with larger areas under water. The authors then show that between 1997 and 2010, partitioned ethnic homelands had a 30% higher likelihood of experiencing political violence (e.g., battles between government forces, rebel groups, and militias) and a 40% higher likelihood of experiencing violence against civilians (e.g., murders, abductions, and child-soldiering raids) than non-partitioned ethnic homelands. This higher incidence of violence is also associated with worse economic outcomes and lower provision of public goods.

4.2.3 Persistent Effects of Historical Institutions

The main conclusion from this set of studies is that dysfunctional institutions are a major obstacle to economic development because they deter investment, create macroeconomic instability, and potentially lead to violent conflict. A crucial question for policy is then whether replacing such institutions reverses these adverse effects. Several papers study this question using natural experiments surrounding the imposition and subsequent abolition of historical institutions.

[Banerjee and Iyer \(2005\)](#) examine the long-term impact of colonial land revenue systems in British India. In some Indian districts, British officials levied taxes on agricultural income, whereas in other districts, the collection of taxes was left to a class of native landlords who were free to set the revenue terms for the peasants under their rule and to dispossess them if they did not pay their dues. After Indian independence in 1947, all direct taxes on agricultural income (and thus both institutions) were abolished. Neverthe-

⁴²Most macroeconomic models imply that macroeconomic volatility reduces growth; see [Scott R. Baker and Nicholas Bloom \(2013\)](#) for estimates of this relationship.

less, Banerjee and Iyer find that to the present day, districts that used the landlord-based system have lower agricultural yields and investment, lower public investment in health and education, as well as worse health and educational outcomes. To show that these relationships are causal, the authors instrument the choice of land revenue system with a dummy variable equal to 1 if the district was annexed between 1820 and 1856, a period during which the authors argue the British preferred the non-landlord-based systems for political and intellectual reasons that are orthogonal to the characteristics of the annexed districts. Banerjee and Iyer argue that these adverse effects of the landlord-based system persist because of its effect on the social structure of the affected areas: it created a class-based antagonism that limits the capacity for collective action in the affected districts to the present day. As a result, these districts are relatively less able to muster the political influence to claim their fair share of expenditure in education, health care, and public goods.

Rather than focusing on heterogeneous institutions within districts that were directly controlled by the British, Iyer (2010) studies the long-term effects of direct British colonial rule versus indirect rule across 415 districts in present-day India. Although all of India was under British political control by the middle of the 19th century, the British administered directly only part of this area (“British India”), while Indian Kings (“Princely States”) who had considerable autonomy ruled the remaining parts. After the end of colonial rule in 1947, all of India then came under a uniform administrative and political structure. The major problem when attempting to estimate the causal effect of direct British control is that the British did not randomly annex areas, but presumably were more eager to annex richer than poorer areas. Iyer deals with this selection problem by using the “Doctrine of Lapse,” a policy that was in place from 1848 to 1856, under which the British annexed princely states whose rulers died without a natural heir. Using lapse as an instrument for direct British rule, she finds no effect on agricultural productivity and investment, but a negative effect of direct British rule on the provision of public goods, education, and health care that lasts through the 1990s, although these persistent effects appear to be decreasing over time.

Whereas both Banerjee and Iyer (2005) and Iyer (2010) rely on instruments in their identification strategy, Dell (2010) uses a regression discontinuity design to study the long-term effects of the *mita* forced labor system that the Spanish instituted in Peru and Bolivia from 1573 to 1812. Under this system, villages within a geographically precisely determined area were required to send one seventh of their male adult population to work in silver and mercury mines. Similar to Pinkovskiy (2013), she argues that the exact

location of the mita boundary is quasi-random within a 50km band, such that all other relevant influences on household consumption vary smoothly at the border of the mita area. Subject to this assumption, Dell estimates that the mita had a persistent negative effect that lowers household consumption by 25% almost 200 years after its abolition. She argues that this effect most likely transmits itself again through the lower provision of public goods and education services to these areas.

Alesina and Fuchs-Schündeln (2007) analyze the long-term effect of institutions on individual economic preferences, as opposed to economic outcomes. They explore German separation and reunification as a natural experiment to analyze how 45 years of communist rule affected individuals' attitudes toward market capitalism, and the role of the state in providing insurance and redistribution from the rich to the poor. Seven years after reunification, East Germans are still much more likely to favor a strong role of the government. This difference in preferences is larger for older cohorts, who lived under different systems for a longer time period. The micro data allow the authors to include rich controls for individual economic motives to favor a strong government, such that the difference can confidently be attributed to the effect of living under a market system versus a communist system. Similar to the convergence found in Iyer (2010), the longer East Germans live under the new market system, the more their preferences resemble those of West Germans. Under the assumption of linear convergence, full convergence will take one to two generations. This convergence is the product of two forces: around one third of the convergence is due to a shift in the cohort composition towards younger cohorts, and around two thirds can be attributed to actual convergence of preferences within individuals.⁴³

The main conclusion from this set of natural experiments is that historical institutions have long-lasting effects—even after their abolition—that transmit themselves through their effect on the distribution of political power, social structure, preferences, or some other mechanism. Although all three studies focusing on economic outcomes commit considerable effort to narrow down the potential channels of persistence, their design does not allow them to identify the exact channel. A remaining challenge for future work is thus to go beyond causal identification of the treatment effect of historical institutions and to identify natural experiments that speak simultaneously to the treatment effect and its channel of persistence.

⁴³Fuchs-Schündeln and Schündeln (2015) similarly find evidence for the endogeneity of democratic preferences.

4.2.4 Determinants and Dynamics of Institutions

If institutions have a large causal effect on economic growth, a crucial question is how countries acquire well-functioning institutions and what might determine their evolution over time. The natural experiments covered so far suggest colonization has generated large and persistent differences in the “level” of the quality of governance and institutions across countries. This view is also consistent with a large literature on the economic consequences of legal origins, which has shown that the identity of the colonizing country determines the type of legal system used in former colonies as well as its performance along a broad range of dimensions (La Porta, Lopez-de-Silanes, Shleifer, and Vishny, 1998, 1997; La Porta, Lopez-de-Silanes, and Shleifer, 2008). Another example of institutional change resulting from foreign domination and conquest is the abolition of feudalism and the imposition of French civil law in German states conquered during the Napoleonic wars (Acemoglu, Cantoni, Johnson, and Robinson, 2011).

Aside from these large-scale cross-sectional experiments, a more practical question, at least from a policy perspective, may be what determines the dynamics of institutions over time. To address this question, a large set of studies attempts to identify exogenous shocks to the political balance of power between different groups within a polity, and how these shocks may foster or retard the development of good political and economic institutions. A closely related set of studies, which we discuss in section 4.3.3, focuses instead on identifying exogenous shocks to a society’s social structure.

Shocks to the Political Balance of Power

One interesting approach to identifying exogenous shocks to the political balance of power uses shocks to the natural environment as an instrument. For example, Brueckner and Ciccone (2011) combine information about GDP per capita and average annual rainfall with measures of democratization and constraints on the executive in sub-Saharan African countries 1981 to 2006.⁴⁴ Because many sub-Saharan countries rely heavily on agriculture, negative rainfall shocks (droughts) may serve as a good instrument for recessions in these countries. Brueckner and Ciccone’s main specification relates changes in constraints on the executive to lagged GDP per capita and a full set of time and country fixed effects, while using lagged average rainfall as an instrument for lagged GDP per capita. They find a negative and highly significant coefficient on GDP per capita. Because

⁴⁴Constraints on the executive typically refer to the extent of institutional constraints on the decision-making powers of the chief executive, such as the president.

the specification uses only variation within countries (due to the inclusion of country fixed effects), this finding implies that transitory recessions are associated with democratization, a result that is in line with a literature that has argued theoretically that autocracies tend to become vulnerable at times of economic crisis (e.g., [Lipset \(1959\)](#), [Huntington \(1991\)](#), [Acemoglu and Robinson \(2005\)](#)).

The main advantage of using changes in the natural environment rather than “man-made” historical events as an instrument is that ruling out reverse causation between GDP per capita and droughts is easy. The main disadvantage is that droughts, and shocks to the natural environment in general, may affect democratization (and many other things) through a variety of channels. Brueckner and Ciccone have two main responses to this concern. First, their standard specification relates changes in constraints on the executive to last year’s GDP growth, such that the timing of the effects mitigates the possibility that droughts may affect GDP through institutions rather than the other way around. Second, they show that the reduced-form effect of rainfall on democratization is much smaller in countries that rely less on agriculture.⁴⁵

[Chaney \(2013\)](#) applies a similar empirical strategy to a completely different context. He argues that in medieval Egypt, the political power of religious leaders tended to increase during years with deviant Nile floods, because economic crises increased their capacity to coordinate a revolt. His main specification shows that in years with deviant floods, Egypt’s main religious figure was less likely to be replaced and that those years also showed more evidence of popular unrest. He then shows evidence from a variety of sources to distinguish his interpretation from a variety of other channels through which deviant floods might have affected Egyptian society.

Rather than using historical variation in floods over time, [Hornbeck and Naidu \(2014\)](#) study the effect of a single event, the Great Mississippi Flood of 1927, on the balance of power between black laborers and white landowners in the affected areas. In 1927, the Mississippi River broke its banks in an unprecedented flood that inundated 26,000 square miles and displaced the affected population. Hornbeck and Naidu argue that this event represented a significant shock to oppressive racial institutions that were geared towards keeping black laborers on the land and in jobs with depressed wages. Subject to the identifying assumption that flooded and non-flooded areas in the same state and

⁴⁵Also see [Miguel, Satyanath, and Sergenti \(2004\)](#) and [Franck \(2015\)](#) for similar natural experiments. Other empirical papers studying the relationship between income and political institutions include [Barro \(1999\)](#), [Glaeser, Ponzetto, and Shleifer \(2007\)](#), [Acemoglu, Johnson, Robinson, and Yared \(2005\)](#), [Persson and Tabellini \(2009\)](#), [Burke and Leigh \(2010\)](#), and [Acemoglu, Naidu, Restrepo, and Robinson \(2014\)](#)

with similar pre-flood characteristics would have developed similarly absent the flood, they show that flooded areas experienced an immediate and persistent out-migration of black laborers. In the following decades, these areas, deprived of cheap labor, modernized agricultural production and increased capital intensity relative to non-flooded areas.

Instead of shocks to a country's natural environment, a large literature on the "resource curse" studies the effect of exogenous changes in the value of countries' endowment of commodities on the quality of their institutions. For example, [Caselli and Tesei \(2015\)](#) calculate the flow of resource rents accruing to commodity-exporting countries. They argue that most commodity exporters are small relative to the world market, such that, for example, a change in the world price of oil is exogenous to Venezuela's political system. Their main specification relates the one-year change in an index of the quality of a country's political institutions to the lagged, three-year average change in the price of its principle export commodity. Their findings are consistent with the idea that an increase in the value of a non-democratic country's natural resources tips the balance of political power in favor of the ruling elite: when autocratic countries receive a positive shock to their flow of resource rents, they tend to respond by becoming more autocratic subsequently.⁴⁶

Popular Mobilization

The main conclusion from the set of papers studying the effect of shocks to a country's political balance of power is that some of these shocks may make it easier to place constraints on the ruling elite, and thus obtain "good" institutions. However, *how* such constraints may be imposed in practice remains unclear, particularly in non-democratic societies that lack a functioning mechanism for replacing the ruling elite. Part of the reason for this lack of evidence is that typical measures of institutional quality vary only at the annual or lower frequency and are thus hard to tie to particular events. In a recent study, [Acemoglu, Hassan, and Tahoun \(2014\)](#) suggest using daily financial data to measure the real-time effects of popular mobilization in street protests on investors' expectations of economic rents from future favoritism and corruption accruing to politically connected firms. Their approach generalizes the event study methodology typically used in the finance literature that estimates the value of political connections from changes in the relative stock market valuations of politically connected firms ([Roberts, 1990](#); [Fisman, 2001](#)).

⁴⁶Also see [Brueckner, Ciccone, and Tesei \(2012\)](#) for a similar exercise.

During Egypt’s Arab Spring, street protests first brought down Hosni Mubarak’s government and then ushered in an era of competition between three groups that repeatedly rotated in and out of power: elites associated with Mubarak’s National Democratic Party (NDP), the military, and the Islamist Muslim Brotherhood. In their main specification, Acemoglu, Hassan, and Tahoun estimate the direct effect of street protests on the valuation of firms connected to the group currently in power relative to their effect on the valuation of non-connected firms:

$$R_{i,t} = I'_{i,t}\beta + (P_t \times I'_{i,t}) \gamma + X'_i \times \delta_t + \delta_t + \eta_s + \epsilon_{i,t}, \quad (16)$$

where $R_{i,t}$ is the log return of firm i on day t and P_t denotes the daily number of protesters in Tahrir Square estimated from Egyptian and international print and online media. $I_{i,t}$ denotes a vector of two dummies, the first reflecting political connections to the group currently in government and the second recording connections to the two other rival (non-incumbent) power groups. X_i is a vector of firm-level controls, and δ_t and η_s denote, respectively, time and sector dummies. The coefficients of interest are the entries of the vector γ , and measure the effect of the number of protesters in Tahrir Square on the relative stock market valuation of firms connected to the incumbent group and the relative valuation of firms connected to the two rival (currently non-incumbent) groups.

The estimates of γ show a robust and quantitatively large negative effect of larger protests on the returns of firms connected to the incumbent group, but no effect on the valuation of their rivals. For example, a turnout of 500,000 protesters in Tahrir Square lowers the market valuation of firms connected to the incumbent group by 0.8% relative to non-connected firms, but triggers no offsetting gain in the value of “rival” (non-incumbent) connected groups. These results hold even when excluding periods from the analysis during which protests may have resulted in a change in regime or any kind of formal institutions.

The main identifying assumption for a causal interpretation of the effect of protests on the relative valuation of firms connected to the incumbent group and their rivals is that (i) no omitted variables should exist that fluctuate at the daily frequency and are correlated with both stock returns and the number of protesters in Tahrir Square, and (ii) no reverse causality should result from daily differential returns on firms connected to different power groups to the intensity of protests. Key to corroborating this assumption is the daily frequency of the data and the timing of the effect: if both stock market valuations and protests respond to some other slow-moving variable, then one might expect stock

returns to be correlated with future protests (the lead of protests should be statistically significant). The authors show that this is not the case in the data. Instead, stock returns respond only during and immediately after the protest. According to their preferred interpretation, these results provide evidence that popular mobilization and protests have a role in restricting the ability of connected firms to capture excess rents and thus may limit favoritism and corruption even if they do not result in changes in formal institutions or the identity of the government.

This interpretation of a causal link between popular mobilization and corruption is also in line with evidence from a natural experiment that links popular mobilization to a range of other political outcomes in the United States: [Madestam, Shoag, Veuger, and Yanagizawa-Drott \(2013\)](#) study the effect of the first large rallies by the Tea Party on April 15, 2009 (the day income tax filings became due). In many regions, these rallies were the first large event organized by the movement. Using rainfall on that day as an instrument, the authors show that higher attendance at these initial rallies (due to good weather) is associated with higher subsequent political support for the Tea Party’s political positions, more votes for the Republican party in the midterm elections of 2010, and more conservative voting records of congressmen subsequently elected in the region.

4.3 Social Structure

A powerful idea in economic sociology is that the economic success of an entity, be it an individual, a household, or a geographic region, depends on its position in the social structure of the marketplace. Well-connected individuals who bridge “holes” in this social structure are more likely to be economically successful and may generate a competitive advantage for the firms at which they work and the regions in which they live ([Loury, 1977](#); [Burt, 1992](#); [Granovetter, 1985, 2005](#)). For example, one might imagine that well-connected individuals provide “social” collateral for economic transactions that would not otherwise be feasible, or reduce informational frictions by providing a credible channel for communication ([Coleman, 1988](#); [Greif, 1993a](#); [Stiglitz, 1990](#)). According to this view, social ties that are formed and maintained for historical or personal reasons are a fundamental cause of economic growth and thus affect the economic development of entire geographic regions.

[Saxenian \(1999\)](#) gives an example of this view. She analyzes the biographies of Indian engineers who migrated to California in the 1970s. Following the liberalization of the Indian economy in 1991, these immigrants were in a position to leverage their social ties to

relatives and friends in Hyderabad and Bangalore. Many excelled in their personal careers, managing outsourcing operations for US firms. Saxenian argues that by connecting Silicon Valley firms to low-cost and high-quality labor in their regions of origin, these Indian immigrants became instrumental in the emergence of their home regions as major hubs of the global IT services industry.

Although empirical evidence of the effect of social ties on a range of microeconomic outcomes is compelling, estimating the effect of social ties on economic growth poses additional difficulties.⁴⁷ The reason is that social ties are likely to be non-random across as well as within regions. Individuals who have social ties may have common unobserved characteristics, sort endogenously across regions, or form social ties in anticipation of future economic benefits (Manski, 1993; Glaeser, Laibson, and Sacerdote, 2002). Identifying a causal link between social ties and macroeconomic outcomes thus requires exogenous variation both in (a) the economic value of social ties and (b) in the formation of these ties across geographic regions. In general, and in the Indian example above in particular, such exogenous variation either does not exist or cannot be measured.

In this section, we first discuss a natural experiment that attempts to address this identification problem, and establish a causal link between social ties and growth. We then turn to the effect of social ties on other aggregates such as trade and foreign direct investment. Compared to the large body of literature on institutions, the study of this fundamental cause of economic growth is in its infancy. In particular, we are unaware of natural experiments that study the dynamics of the formation and value of social ties over time.

4.3.1 The Effect of Social Ties on Growth

Burchardi and Hassan (2013) use the natural experiment surrounding the fall of the Berlin Wall to identify the causal effect of social ties on economic growth. They argue this setting's key advantage is that the partition of Germany was generally believed to be permanent. After the physical separation of the two German states in 1961, private economic exchange between the two Germanies was impossible. As a result, West Germans who maintained social ties with East Germans during this period did so for purely non-economic reasons. After the fall of the Berlin Wall, trade between the two Germanies

⁴⁷These microeconomic outcomes range from education (Sacerdote (2001)) and employment (Munshi (2003)) to performance in the financial industry (Cohen et al. (2008)) and agricultural yields (Conley and Udry (2010)). Also see Bertrand et al. (2000), Hochberg, Ljungqvist, and Lu (2007), Beaman (2012), Kuhnen (2009), Shue (2013), and Banerjee, Chandrasekhar, Duflo, and Jackson (2013).

suddenly became feasible. To the extent that social ties facilitate economic exchange, the fall of the Berlin Wall thus generates exogenous variation in the value of these ties (condition (a) stated above).

In their main specification, Burchardi and Hassan study the growth in income per capita across West German regions in the six years following the fall of the Berlin Wall as a function of the share of the region's population that has ties to relatives in East Germany. Their structural equation of interest is

$$Y_r^{95} - Y_r^{89} = \beta T_r^{89} + Z_r' \gamma + \varepsilon_r, \quad (17)$$

where Y_r^t is log income per capita in region r in year t , T_r^{89} is a proxy for the share of the region's population that has ties to the East, and Z_r is a vector of controls that contains a complete set of federal state fixed effects, log income per capita in 1989 (Y_r^{89}), the growth rate of income per capita between 1985 and 1989, and the distance from region r to the inner-German border. Because the specification controls for the pre-trend in growth, the coefficient β estimates the *differential change* in the growth rate of income per capita after 1989 for regions with different intensities of social ties to the East.

Equation (17) consistently estimates the parameter of interest if $Cov(T_r^{89}, \varepsilon_r) = 0$. However, this covariance restriction may not hold in the data, because the strength of social ties to East Germany may be correlated with differences in growth prospects across regions. An unbiased estimate of β thus requires exogenous variation in the regional distribution of social ties (condition (b) stated above). Burchardi and Hassan argue that such variation arises as the result of a large-scale migration post World War II when millions fled from East Germany to the West. An overwhelming concern for the migrants arriving in the West at the time was an acute lack of housing. During World War II, almost a third of the West German housing stock was destroyed. Variation in wartime destruction thus made settling in some parts of West Germany more difficult than in others at a time when millions of migrants were arriving from the East. Burchardi and Hassan argue that the extent of regional destruction in 1946 thus provides the exogenous source of variation in the regional distribution of social ties needed for identifying β .

Their key identifying assumption combines a difference in differences approach with instrumental variables: it states that, conditional on the covariates in Z , (i) no omitted variable drives both wartime destruction and differential changes in income growth post 1989, and (ii) wartime destruction in 1946 has no effect on changes in the growth rate of income per capita after 1989 other than through its effect on the settlement of migrants

who have social ties to the East.

Using the degree of wartime destruction as an instrument for the share of the population with social ties to the East, Burchardi and Hassan then explicitly test the hypothesis that a concentration of households with social ties to East Germany in 1989 in a given West German region is causally related to a rise in the growth rate of income per capita after the fall of the Berlin Wall. They find that regions that received a one-standard-deviation larger share of migrants from the East prior to 1961 experienced a 4.6-percentage-point higher growth rate of income per capita in the six years following the fall of the Berlin Wall. Although both entrepreneurs and non-entrepreneurs who live in regions with strong social ties to the East experience a rise in their incomes post 1989, the incomes of entrepreneurs increase at a significantly higher rate. Moreover, the share of the population engaged in entrepreneurial activity rises in regions with strong social ties to the East. Consistent with this increase in entrepreneurial activity, West German firms that are headquartered in regions with strong social ties to the East are more likely to invest in East Germany between 1989 and 2007.

A crucial question for the interpretation of these results is whether this link between social ties and growth is a purely “microeconomic” phenomenon in the sense that a few people who have ties to the East internalize all of the benefits from this tie, or whether it is a “macroeconomic” phenomenon that involves positive spill-overs from one person’s tie to the East to the income growth of un-connected individuals living in the same region.

Using household-level data, Burchardi and Hassan show that the income growth of households with ties to at least one relative in the East is on average 6 percentage points higher in the six years following the fall of the Berlin Wall than that of comparable households with no such ties. The authors then relate the effects of social ties on household and region-level income growth using a model in which household income is a function of direct and indirect (higher-order) social ties to the East, where individual households may benefit from having friends with ties to the East, even if they themselves have no such ties. Their preferred estimates imply that, other things equal, a direct social tie to the East has the same effect on individual household income as a 50-percentage-point (or 3.5-standard-deviation) increase in the regional share of households with such ties. From the perspective of an individual household, the incremental benefit from a direct social tie to the East is thus large compared to the incremental benefit from higher-order social interaction. Nevertheless, indirect social ties to the East account for two thirds of the aggregate effect, because all of a region’s households benefit from indirect social ties, whereas only a subset of the population benefits from direct social ties.

4.3.2 The Effect of Social Ties on Trade and FDI

Aside from an interest in identifying the fundamental causes of economic growth, trade economists have long documented a strong association between social ties and the pattern of international trade and foreign direct investment (FDI) as part of a broader effort to understand the role of informal barriers in shaping economic interactions across borders (Gould, 1994).⁴⁸ The central puzzle in this literature is that geographic distance and country borders have a strong negative impact on bilateral trade flows, even after controlling for any measurable barrier to trade. Part of this puzzle may be resolved if social ties are effective means of overcoming informal barriers to trade (e.g., informational frictions or problems with contract enforceability) and are negatively correlated with geographic distance and country borders.

Perhaps the most influential of these studies is Rauch and Trindade (2002), who study the effect of ethnic Chinese networks on international trade. They study Chinese networks in particular because data on the population share of ethnic Chinese are readily available and because Chinese migrants are present in most countries of the world. In their main specification, the authors estimate a conventional gravity equation that explains bilateral trade as a function of the product of the GDP of the two countries, geographic distance, and the product of the share of the population of the two countries that is ethnic Chinese. This product can be interpreted as a proxy for the strength of social ties between Chinese residents in the two countries – the probability that if one selects at random an individual in each country, both will be ethnic Chinese. Focusing on countries in which ethnic Chinese constitute at least 1% of the population (e.g., as in all of Southeast Asia), this variable has remarkable predictive power for bilateral trade, suggesting trade would be on average around 60% lower in the absence of the Chinese ethnic network. This conditional correlation is stronger for trade in differentiated goods, which is consistent with the view that the Chinese ethnic network facilitates trade by reducing asymmetric information. Combes, Lafourcade, and Mayer (2005) find a similarly strong conditional correlation between the volume of trade between a given pair of French regions and the bilateral stock of migrants between the two regions, where they measure the stock of migrants as the number of individuals born in region i that work in region j . Once they add this variable to their gravity equation, the estimated effects of distance and region borders on trade are reduced to a much more reasonable level (both effects drop by about 50%).

⁴⁸A number of other papers show that measures of affinity between regions, such as trust, telephone volume, and patterns of historical migration, correlate strongly with other aggregate outcomes, such as foreign direct investment (Guiso et al., 2008b) and international asset flows (Portes and Rey, 2005).

Garmendia, Llano, Minondo, and Requena (2012) find similar results for Spain.

The main conclusion from this set of papers is that failing to account for the effect of social ties across borders and within countries may explain why traditional gravity equations have found unreasonably large effects of distance and country borders on trade. However, none of the three studies attempt to deal with potential endogeneity or reverse causality, such that the conditional correlations between trade flows and social ties that they document should not be interpreted as causal effects. In other words, the effect of social ties on trade flows is not identified in these papers for the same reason that (17) is not identified without an appropriate instrument.

Three recent studies address this issue using different natural experiments. Parsons and Vezina (2014), Cohen, Gurun, and Malloy (2014), and Burchardi, Chaney, and Hassan (2015) evaluate the causal impact of migrant networks on the variation in international trade and FDI across locations in the United States. To the extent that these papers look at trade and FDI originating from the same country, the United States, some of the concerns of reverse causality are mitigated: the regulatory environment, most direct barriers to trade and investment, as well as the ease with which migrants from different countries can emigrate are all relatively uniform.

Parsons and Vezina's strategy is similar to that of Burchardi and Hassan (2013) in the sense that they study variation in social ties that results from a historical migration that occurred while economic interaction with the migrants' region of origin was impossible. At the end of the Vietnam war, the US government imposed a trade embargo on Vietnam and evacuated 130,000 Vietnamese citizens to the United States. Upon arrival at one of four processing centers located in Arkansas, California, Pennsylvania, and Florida, charitable organizations were charged with finding sponsors who were willing to provide food and shelter for the refugees. Parsons and Vezina stress that a main objective of this process was to disperse the Vietnamese refugees as much as possible across the United States to avoid an agglomeration of refugees, similar to that of Cubans in Florida. Consistent with this view, they show that the resulting allocation of refugees is uncorrelated with a range of variables that may proxy for the potential for trade with Vietnam, which is quite plausible because of the trade embargo that was in force at the time the refugees were allocated. Instead, they argue that the variation in the allocation of refugees across states was driven by quasi-random variation in the capacity of the charitable organizations operating in different states. Importantly, Parsons and Vezina also show that the initial allocation of this first wave of refugees in 1975 is highly predictive of the location of ethnic Vietnamese in 1995, the year in which the trade embargo was finally lifted. Their

structural equation of interest takes the form

$$X_i = \beta V_i + C_i' \gamma + \epsilon_i, \quad (18)$$

where X_i is the average share of exports of state i to Vietnam between 1995-2010, V_i is the stock of Vietnamese migrants in 1995, and C_i is a vector of controls. The standard specification uses the allocation of refugees across the United States in 1975 as an instrument for V_i . The identifying assumption for a causal interpretation of β is thus that the initial allocation of refugees is uncorrelated with ϵ_i . Subject to this assumption, Parsons and Vezina's main specification implies that a doubling of the population share of Vietnamese migrants relative to the mean increases the ratio of exports to Vietnam over GDP by 19.8%.

Cohen, Gurun, and Malloy (2014) instead use the forced relocation of ethnic Japanese into Japanese Internment Camps during World War II as an exogenous shock to the location of ethnic Japanese across US Metropolitan Statistical Areas (MSAs). These camps were established in remote areas (away from any industrial activity that may be considered sensitive to the war effort) to house Japanese-Americans who lived predominantly on the West Coast prior to the bombing of Pearl Harbour. After the war, the residents were released from the camps. However, having lost their jobs and sold off their possessions in their regions of origin, they often re-settled in MSAs that were geographically close to the location of their internment. Cohen et al. (2014) show that this relocation had a persistent effect on the regional distribution of the Japanese-American population that lasts to the present day: MSAs that are within a 250-mile radius of the location of a former internment camp have a 62% larger Japanese population today than other comparable MSAs. To corroborate this finding, they show, importantly, that interment camps predict higher populations of Japanese-Americans but not of other Asian-Americans.

This persistent effect on the location of Japanese-Americans has a sizable effect on the probability that firms located in a given MSA trade with Japan. Using the location of Japanese internment camps as a instrument, they show that a one-standard-deviation increase in the share of an MSA's population that are Japanese-Americans doubles the likelihood that a given firm will export to Japan. They find similarly large effects on the probability of importing, the volume of trade, and even on the likelihood that a given MSA has a sister city in Japan today.

Although both Parsons and Vezina (2014) and Cohen et al. (2014) make convincing arguments for a causal link between social ties and trade, both use natural experiments

that resulted in a shock to the allocation of migrants from one particular country, such that they cannot control for unobserved destination effects as one would in a gravity equation. Assessing the external validity of these results and how they generalize to other ethnicities is therefore difficult.

To quantify the more general causal effect of social ties on US trade and FDI, [Burchardi et al. \(2015\)](#) instead study the natural experiment that arises from the entire history of settlement of the United States. Their strategy isolates quasi-random variation in the allocation of migrants across destinations within the United States that results from the interaction of two facts: First, migrants from different origins tended to arrive in the United States at different times. Second, the set of destinations that are most economically attractive to the typical migrant arriving in the United States changed over time.

They motivate their approach using a reduced-form dynamic model of migrations. Migrations from a given foreign origin country o to a given destination county d in period t depend on the total number of migrants arriving in the United States from o (a push factor), the relative economic attractiveness of destination county d to migrants arriving at the time (a pull factor), and the size of the pre-existing local population of ancestry o , allowing for the fact that migrants tend to prefer settling near others of their own ethnicity (a recursive factor). Solving the model shows the number of residents today who are descendants of migrants from o is a function of simple and higher-order interactions of the sequence of pull and push factors. [Burchardi et al. \(2015\)](#) then use these interactions to construct instruments for each US county's present-day ancestry composition.

To prevent omitted variables that affect both migrations and FDI from driving their results, they measure the pull factor of each US destination for migrants from o , using the number migrants arriving in d at the same time from a continent other than o 's continent of origin. That is, they predict a migrant's choice of destination within the United States using the revealed behavior of the average migrant arriving at the same time but from a different continent. Similarly, they measure the push factor using the total number of migrants arriving in the United States from o at time t , excluding those who settle in the vicinity of d . Interacting these measures of pull and push factors for each vintage of census data since 1880 isolates variation in the present-day ancestry composition of US counties that derives solely from the interaction of the staggered arrival of migrants from different origins with time-series variations in the relative attractiveness of different destinations within the United States.

In their main specification, Burchardi, Chaney, and Hassan find that doubling the number of individuals with ancestry from a given foreign country increases by 4.2 per-

centage points (or 237% relative to the mean) the probability that at least one firm from that US county engages in FDI with that country. The main identifying assumption for a causal interpretation of this result is that omitted variables making a given location within the United States differentially more attractive for migrants from a specific origin for both settlement and FDI do not affect the location choices of the average migrant originating from other continents and simultaneously have large effects outside of the surrounding states of the destination in question.

Flexibly applying this instrumentation strategy to the entire set of origins and destinations, the authors estimate heterogeneous effects of ancestry on FDI across origins and destinations. Across origins, they find that the effect increases with the geographic distance and the judicial quality of the origin country. Across destinations, they find consistent evidence of a positive impact of regional diversity on FDI: the more diverse and less ethnically homogeneous the local population, the larger the total effect of ancestry on FDI. Similarly, the effect of ancestry on FDI falls with the population of migrants from the same origin in neighboring counties, and from neighboring origins, so that a small minority from a distant part of the world that is not otherwise represented in the local ethnic mix has the largest marginal impact on FDI.

By instrumenting separately for each wave of immigration, [Burchardi et al. \(2015\)](#) are also able to distinguish the effect of first-generation immigrants from the effect of their descendants. They find a significantly smaller effect for the first generation, implying the full effect of ancestry on FDI is long lasting and takes multiple generations to fully unfold.

Simultaneously instrumenting for migration from multiple origins also enables the authors to base their estimation on a gravity equation, including destination fixed effects that control for differences in size, market access, and productivity across US destinations. Although the results with respect to FDI remain stable regardless of the inclusion of destination and other (more complicated) fixed effects, the authors find no systematic causal impact of ancestry on the patterns of international trade of US states whenever they control for destination fixed effects. This latter finding is in stark contrast to the results in [Parsons and Vezina \(2014\)](#), and [Cohen et al. \(2014\)](#), and suggest that the causal relationship between social ties and trade found in these papers may not generalize to a larger set of origin countries.

4.3.3 The Effect of Internal Social Structure on Institutions and Growth

Although the literature on social ties focuses on the external relationships of a social entity, a number of studies have also used natural experiments to assess the effect of the *internal* relationships between different groups within a given society on aggregate economic outcomes, and in particular on the ability of a society to develop a functional political system and “good” institutions.

Perhaps the most convincing of these studies is [Dippel \(2014\)](#), who considers the natural experiment surrounding the formation of Native American reservations. In the 19th century, the US government formed several reservations consisting of members of ethnically and linguistically homogeneous tribal bands. While respecting ethnic differences, this process largely ignored differences in historical institutions, such that some (“mixed”) reservations received constituents of several previously politically independent sub-tribal bands while others did not. Dippel shows that these mixed reservations have significantly worse contemporary economic outcomes, even when conditioning only on variation within reservations belonging to the same tribe. To account for potential confounding factors in the formation of reservations, he instruments the likelihood of a mixing of several previously independent bands with historical mining activity in the historic homeland of the tribe, where mining activity generated incentives for the US government to form fewer, and thus more likely mixed, reservations. He then shows that the majority of the divergence in economic outcomes appears only after the 1980s, when the Bureau of Indian Affairs ceded reservation governance to the local reservations. Using information on contemporary political conflict and corruption within reservations, he argues convincingly that the adverse economic effects are explained by the fact that mixed reservations tend to have more dysfunctional political institutions.

Another set of studies has linked the emergence of good institutions to the relative influence of the middle class. [Acemoglu, Johnson, and Robinson \(2005\)](#) use a difference-in-differences approach to argue that Western European countries that developed a large and politically influential merchant class were able to develop constraints on the executive and safe property rights that were crucial for subsequent economic growth. Their evidence comes from a panel data set of GDP per capita, institutional quality, and the number of Atlantic voyages undertaken by each European country 1300-1850. They show that the onset of the Atlantic trade after 1500 was a major positive shock to GDP per capita in countries on the Atlantic coast. However, this shock led to an increase in constraints on the executive only in Atlantic countries that already had relatively higher constraints

on the executive in medieval times: the interaction between Atlantic trade and medieval institutions explains most of the variation in institutional quality across European countries. The authors argue that these results are consistent with their view that a merchant class could only emerge in countries in which rulers were relatively constrained and did not monopolize the Atlantic trade.

Although AJR (2005) provide historical and anecdotal evidence to corroborate their interpretation, a problem with their analysis is that they do not observe the size of the merchant class directly. More generally, a major challenge for the literature attempting to establish a causal link between social structure and institutions is that consistent measures of social structure are rarely available for a sufficiently long period of time, and in particular for historical episodes in which one might observe quasi-exogenous variation in social structure.

While stopping short of claiming success at having identified a causal effect, [Acemoglu, Hassan, and Robinson \(2011\)](#) make some progress in this dimension by studying the mass-murder of Jews following the Nazi invasion of Russia during World War II. Uniquely, Soviet authorities kept extensive records of the size and ethnic composition of the middle class (“white-collar workers”) in Russian regions (oblasts), dating back to 1926. Before the outbreak of World War II, Jews were heavily over-represented in white-collar occupations, such that their persecution and murder by the occupying forces represented a significant shock to the size of the middle class. Using variation both within oblasts occupied by the Nazis and across occupied and non-occupied oblasts, the authors show that oblasts in which the Holocaust most severely reduced the size of the middle class have worse political and economic outcomes today. These oblasts grew less since 1945 both in terms of population and GDP per capita, exhibited greater vote shares for communist candidates during the 1990s, and more support for preserving the Soviet Union in a referendum held in 1991. Moreover, the shock to the relative size of the middle class in oblasts most adversely affected by the Holocaust appears to persist over time, until the last Soviet census held in 1989. Acemoglu, Hassan, and Robinson argue that their evidence is consistent with the view that a shock to the size of the middle class may have permanent effects because it reduces the core constituency for policies that advance constraints on the executive and safe property rights (most notably in the form of more political support for the preservation of communism).

4.4 Trust and Civic Capital

The literature on culture and economics offers various competing definitions of social capital, culture, trust, and related concepts. For the purposes of structuring our discussion, we focus on the concept of “civic capital,” which [Guiso et al. \(2012\)](#) define as “those persistent and shared beliefs and values that help a group overcome the free rider problem in the pursuit of socially valuable activities.” We prefer to use this relatively narrow definition mainly because it allows a convenient grouping of the existing empirical literature. First, it clearly distinguishes civic capital from social structure, the focus of the literature discussed in the previous section. We may loosely think of “social capital” as the union of the two concepts. Second, it closely describes the variables used in the empirical literature, which often focuses on various measures of how much individuals are willing to trust a stranger, and other measures of beliefs about the intentions and actions of others.

The idea that the set of norms and beliefs that make cooperation among individuals easier should be a driver of economic growth has a long tradition in the social sciences, going back at least to [Banfield \(1967\)](#), [Coleman \(1988\)](#), and [Greif \(1993b\)](#). [Putnam, Leonardi, and Nanetti \(1993\)](#) famously argued that Southern Italy is less developed economically than the North because of a lower level of civic capital, and conjectured that this difference is a result of the fact that Northern cities had a long tradition of self-rule, which fostered a tradition of civic engagement, at a time when Southern cities were tightly controlled by Norman kings.

The basic idea in this literature is that culture in general and civic capital in particular changes only slowly over time. Parents pass on beliefs and values to their children, such that civic capital is akin to a slow-moving, accumulated factor. Societies with a higher level of civic capital have a higher capacity for economic growth because they develop better tools for overcoming market failures and collective action problems.

The prime example for such a belief is trust. Even in a society with an efficient police force and functioning courts, most commercial transactions involve some element of trust ([Arrow, 1972](#)). For example, when you hire an accountant to do your taxes, you trust that she will not abuse your private information to commit credit card fraud. If you cannot trust your accountant, you might prefer to do your taxes yourself. Similarly, when you take a taxi, you trust that the driver knows how to drive, is not intoxicated, has not manipulated the meter, and will not simply lock the door, drive you off to the desert, and hold you for ransom. Although you can take accountants and taxi drivers to

court, doing so will cost time and money, and even the most efficient court systems cannot enforce all the rules all the time. Other transactions that rely on trust include employment contracts in which managers cannot perfectly monitor employees, sales in which goods are delivered before or after payment is made, and many financial transactions and investment decisions. (Thinking about it this way, the amount of trust people in developed societies put in perfect strangers is remarkable!) The more complex an economy becomes, and the more labor is divided into specialized tasks, the more important may be the shared belief that strangers can generally be trusted.

4.4.1 The Effect of Trust on Growth

A number of papers have used cross-country data to document a conditional correlation between measures of civic capital and economic development. [Knack and Keefer \(1997\)](#) show that measures of trust and civic cooperation are strongly associated with higher GDP, higher economic growth, and higher investment-to-GDP ratios, even after controlling for education, institutions, and other factors. However, these results do not speak to causation. The obvious problem is that people who live in wealthy countries with good institutions may rationally put more trust in strangers because they know they will be partially protected by a well-functioning police force and efficient courts. Clearly, institutions, civic capital, and economic development are mutually interdependent variables. A key challenge in demonstrating a causal effect of civic capital on growth is thus not only to identify exogenous variation in civic capital, but also to separate its effect from the effect of institutions.

Three papers attempt to tackle this challenge using natural experiments that rely on the idea that civic capital depends on the experiences of each generation and is at least partially transmitted from one generation to the next.⁴⁹ For example, one might expect individuals whose parents grew up under an authoritarian dictatorship to be less trusting than otherwise similar individuals whose parents grew up in a democracy.

[Tabellini \(2010\)](#) uses the fact that some Western European countries emerged as the union of several very heterogeneous historical political entities. He studies the variation in gross value added per capita across 69 regions within these countries. To the extent that present-day institutions vary only at the country-, rather than the regional-level, country fixed effects absorb any differences in current institutions. He then instruments

⁴⁹Several studies show trust indeed has an inherited component; see [Rice and Feldman \(1997\)](#), [Putnam \(2000\)](#), and [Guiso et al. \(2006\)](#).

for current measures of civic capital using literacy in 1880 and a measure of constraints on the executive between 1600 and 1850, while controlling for urbanization in 1850. The key identifying assumption is that these historical instruments affect present-day economic development only through their effect on the persistent component of civic capital. Conditional on this (rather demanding) assumption, Tabellini shows that the exogenous component in civic capital has a large positive effect on regional economic development. A more conservative interpretation of the same fact, that is nevertheless interesting in its own right, is that distant political history is an important determinant of current economic performance not just across but also within countries.

Guiso, Sapienza, and Zingales (2008a) consider a similar experiment within Italy. In the Northern part of Italy, some cities achieved the status of a free city during medieval times, whereas others remained under the control of a feudal lord or the Holy Roman Emperor. Consistent with the results in Tabellini (2010), they find that those cities that achieved self-rule earlier (by 1136 or 1300 CE) exhibit higher measures of social capital today. However, Guiso et al. (2008a) then go one step further by instrumenting the dummy variable for early self-rule with two variables that they argue historically affected the cost of achieving self-rule but are unlikely to affect directly the level of civic capital or the level of output today: whether the city was a seat of a bishop who may have been able to coordinate the struggle for independence, and whether the city was founded in pre-Roman (Etruscan) times and is therefore located in a geographic position that is easy to defend militarily. Using these two instruments, they confirm that a longer history of self-rule is significantly associated with higher social capital and higher GDP per capita today.

Although more robust than Tabellini (2010), the identifying assumption for a causal interpretation of the results in Guiso et al. (2008a) is still a tall order: self-rule, the two variables driving the cost of self-rule, or any omitted variables correlated with these measures cannot have a direct effect on GDP per capita today, except through their effect on civic capital. Even if formal institutions did not vary within Italy (which they do to some extent), administrative capacity, the functioning, and the quality of these institutions surely vary across regions even though they might follow the same letter of the law. As a result, both of these natural experiments may still confound the effects of civic capital and institutions or any other omitted variable that is correlated with these variables.

Two recent papers attempt to tackle this problem using clever strategies that are not natural experiments according to our definition. Algan and Cahuc (2010) use the inherited

trust of descendants of US immigrants covered in the General Social Survey to recover a long time series of trust for their origin countries, reaching back to the beginning of the 20th century. This long time-series then allows them to relate changes in GDP per capita to changes in inherited trust over several generations, while controlling for all of the time-invariant effects that complicate the interpretation of the results in [Tabellini \(2010\)](#) and [Guiso et al. \(2008a\)](#). In addition, they also control for time variation in the quality of institutions in order to convincingly distinguish the effect of institutions from the effect of trust. Their estimated effect of trust is positive, highly statistically significant, and quantitatively large. For example, according to these estimates, GDP per capita in Russia and Mexico would have been 60% higher had these two countries inherited the same level of trust as Swedes. [Gorodnichenko and Roland \(2010\)](#) also rely on the idea that cultural traits are inherited and instrument culture (in their case, a measure of individualism rather than trust) with the genetic distance of the population to the most individualist countries in the world (the United States and UK). Consistent with the other studies, they also find a large effect of culture on growth.

4.4.2 Effect of Trust on Financial Development and Other Aggregates

Although the papers surveyed in the previous section may have convincingly identified a causal effect of civic capital on growth, they have little to say about the mechanism through which this effect transmits itself and about *how* civic capital affects growth. One obvious candidate is financial development: financial contracts are arguably “trust intensive”, in the sense that handing over cash to a stranger today in the hopes of receiving returns in the future requires a large amount of trust in that stranger. Higher levels of civic capital may thus enable a society to sustain a more sophisticated financial system that may then in turn facilitate economic growth. [Guiso, Sapienza, and Zingales \(2004\)](#) study this channel.

Similar to [Guiso et al. \(2008a\)](#), this paper measures variation in civic capital across regions within Italy and relates these measures to the use of financial instruments by households responding to a survey by the Italian central bank. The identification strategy again relies on the idea that part of social capital is inherited from previous generations. Thus, when a household moves from one Italian region to another, the level of civic capital (but not the quality of institutions) in its region of origin still influences its behavior. In their main specification, [Guiso, Sapienza, and Zingales \(GSZ, 2004\)](#) relate a household’s use of financial instruments to the level of civic capital (measured as voter turnout or

the volume of blood donations) in its region of origin, a set of region fixed effects, and a number of household-level controls. They find that households that originate in regions with higher levels of social capital are more likely to use checks, invest more in the stock market, and rely less on informal loans from friends and family. These effects tend to be stronger in regions with weak law enforcement.

The authors' preferred interpretation of these results is that the civic capital plays an important role in the degree of financial development across Italy. If these results generalize to the variation in financial development across countries, they may thus account for part of the observed effect of civic capital on economic growth.

Apart from this evidence of a financial channel, additional evidence on the mechanism by which civic capital affects growth is scarce, and if it exists largely not causally identified. A promising avenue for future applications of natural experiments may be the relationship between trust and regulation. For example, [Aghion, Algan, Cahuc, and Shleifer \(2010\)](#) document that levels of trust and the level of government regulation are strongly negatively correlated across countries.

4.4.3 Determinants and Dynamics of Trust

The main conclusion from the series of studies that document an effect of civic capital on growth and financial development is that, from a macroeconomic perspective, we may think of civic capital as a slow-moving state variable that co-determines a society's capacity for economic growth. A crucial question then is what governs the dynamics of this state variable. That is, how do some societies end up with a high level of civic capital while others suffer from low levels of civic capital? The existing literature has examined natural experiments that identify three factors determining the dynamics of civic capital: historical institutions, experiences of violence and conflict, and the climate.

Historical Institutions

As part of their identification strategies, [Tabellini \(2010\)](#) and [Guiso et al. \(2008a\)](#) show that historical institutions appear to affect the level of trust, decades or even centuries later. Both papers show that the descendants of residents that lived in areas that had more constraints on the executive historically exhibit higher levels of trust today.

[Becker, Boeckh, Hainz, and Woessmann \(2015\)](#) examine a natural experiment in which one might expect similar results. Parts of five Eastern European countries (Montenegro, Poland, Romania, Serbia, and Ukraine) were under the rule of the multi-ethnic Habsburg

empire until the end of World War I. In some parts, this rule lasted for hundreds of years. Compared to both its contemporaries and some of its successor states, the Habsburg empire had a reputation for having a restrained and effective bureaucracy, courts, and police. The descendants of residents of areas formerly under Habsburg control thus had a longer history of living under “good” institutions in this sense than their present-day countrymen. [Becker et al. \(2015\)](#) study the effect of this treatment using a regression-discontinuity design.

Their main specification relates responses from a survey covering individuals in all five countries to their location relative to the former border of the Habsburg empire. They find that individuals living within 200 km of the former Habsburg side of the border are not significantly more trusting of strangers or more likely to be members of a civic organization than their countrymen on the other side of the former border. Although this finding may be due to a lack of power in their specification, they do find that individuals on the former Habsburg side are significantly more trusting of the police and less likely to pay bribes to officials.

History of Violence or Conflict

One interesting detail in the results of [Algan and Cahuc \(2010\)](#) is that inherited trust in Sweden increased after 1935 while it decreased in continental Europe and the UK. They conjecture that this differential change may be the effect of World Wars I and II: individuals that experience violence and conflict may pass down a lower level of trust to their descendants. Two papers examine this link between violence and trust in different historical settings.

[Nunn and Wantchekon \(2011\)](#) link the variation in levels of trust across individuals in Africa to the history of slave trade. They argue that the demand for slaves by Europeans (and later Americans) based predominantly in Atlantic ports created conditions that would result in distrust among the indigenous population. Particularly in the later phases of the slave trades, individuals found themselves enslaved not as the result of inland raids by foreigners but as the result of kidnappings or trickery on the part of other members of the indigenous population. By selling others into slavery, one could obtain the means to purchase iron weapons, thus protecting oneself from enslavement. A number of historical sources show that the majority of individuals were sold into slavery by kidnappers or even by their own relatives. Exposure to the slave trade may therefore have severely reduced the level of trust in the affected societies.

Nunn and Wantchekon’s structural equation of interest takes the following form:

$$trust_{j,e,d,i} = \alpha_i + \beta \textit{slave exports}_e + X'_{j,e,d,i}\Gamma + X'_{d,i}\Omega + X'_e\Phi + \epsilon_{j,e,d,i}, \quad (19)$$

where α_i denotes country fixed effects, *slave exports*_{*e*} measures the number of slaves taken from ethnic group *e* during the slave trade per square kilometer of area settled by the ethnic group, $X'_{j,e,d,i}$ denotes a rich set of individual-level controls including ethnicity, education, and age, $X'_{d,i}$ controls for the ethnic composition of district *d* in country *i*, and X'_e is a vector of ethnicity-level controls that capture subnational variation in colonial rule, in particular, the disease environment and measures of pre-colonial prosperity.

In their main specification, Nunn and Watchekon instrument slave exports with the distance of an ethnic group from the coast, which is where European slave traders kept their bases. Because geographic features in general are correlated with all kinds of things, they then make a careful argument that, conditional on the controls they include in their standard specification, the distance to the coast is plausibly uncorrelated with other factors that affected trust. First, they argue that Africans did not engage in overseas trade before the slave trade, such that distance to overseas trade is not a confounding factor in their case. Second, they stress the importance of the ethnicity-level controls for other forms of European contact. Third, they include additional controls for each ethnicity’s historical reliance on fishing. Conditional on these controls, they argue that the exclusion restriction is plausibly satisfied. Their estimates show that a one-standard-deviation increase in exposure to the slave trade is associated with approximately a 0.2-standard-deviation decrease in various measures of trust of neighbors and trust of other ethnicities.

A fairly common problem with papers using large natural experiments such as the enslavement of Africans is that unobservables could potentially bias the result. For example, in [Nunn and Wantchekon \(2011\)](#), we may worry that despite the large number of controls the authors propose, differences in pre-existing trust and prosperity or some other unobservable correlated with slave trade and trust may not be adequately accounted for. To assuage these concerns, Nunn and Wantchekon use a technique developed by [Altonji, Elder, and Taber \(2005\)](#) and [Bellows and Miguel \(2009\)](#) that calculates how much stronger selection on unobservables, relative to selection on observables, would have to be to overturn the estimated effect:

$$\hat{\beta}^R / (\hat{\beta}^R - \hat{\beta}^F),$$

where $\hat{\beta}^R$ and $\hat{\beta}^F$ are the coefficients of interest estimated with a restricted and the full set of controls, respectively. If including observable covariates does not have a large effect on the coefficient of interest, this number is large and selection on unobservables would have to be multiple times more severe than selection on observables to overturn the result. Nunn and Wantchekon find that including their full set of controls changes their coefficient of interest so little that this statistic is above 3 in all of their specifications. Whatever their specifications are missing would thus have to have a very large selection effect to overturn their qualitative result.

A second paper probes the relationship between inter-state conflict and trust. [Jancec \(2012\)](#) uses a differences-in-differences approach to show that individuals living in regions within the present-day countries of Slovenia, Croatia, Serbia, Montenegro, Romania, and Ukraine that experienced more frequent changes in the ruling nation state between 1450 and 1945 exhibit lower trust in political institutions today. However, similar to [Becker et al. \(2015\)](#), he finds no effect on measures of civic behavior and trust toward strangers.

Geography and Climate

Rather than identifying “man-made” shocks, such as wars and historical migrations, [Durante \(2010\)](#) takes a more radical approach and links civic capital directly to environmental factors that determine the need for cooperation. The idea (similar to [Ostrom \(1990\)](#)) is that the earliest societies, centered around subsistence farming, would develop a culture of cooperation and trust where it is needed for survival. He argues that in regions where precipitation and temperature are highly variable from year to year, societies needed to develop civic capital to sustain investment in irrigation and other large works that facilitated survival. Similarly, in regions with very diverse climatic conditions, developing civic capital increases the probability of survival, by facilitating trade and risk-sharing. Using long-term climate data reaching back to 1500 CE, he indeed finds a significant association between trust and these climatic variables in the cross section of European regions. Interestingly, these results are robust to including country fixed effects and controlling for early institutions as in [Tabellini \(2010\)](#), suggesting that part of the effect of climate on trust may indeed transmit itself through civic capital.

4.5 Multiple Equilibria and Path Dependence

Some growth models that feature non-convexities produce multiple equilibria, such that the mapping between the steady state GDP per capita and ϕ_i in (13) is not unique. For

example, [Murphy, Shleifer, and Vishny \(1989\)](#) show that multiple equilibria can arise in a simple model with monopolistic competition and a fixed cost of production. In their model, a given firm finds that incurring the fixed cost of production is profitable only if other firms do the same, due to a demand spillover. In the “bad” equilibrium, firms do not invest, because they expect that other firms will also not invest. This scenario is an example of a coordination failure – economic growth does not happen because economic actors have the “wrong” expectations and cannot coordinate to invest simultaneously. Although it seems implausible that countries might be persistently poor just because its residents cannot coordinate to simultaneously change their expectations, many more complicated models also feature multiple steady states. In these models, long-run GDP per capita is path-dependent, and once an economy finds itself on the path to a bad steady state, it may be hard to reverse its trajectory.

To our knowledge, no single natural experiment has been used to test the hypothesis that multiple steady states may explain cross-country income differentials. Part of the reason for the absence of such an experiment might be that viewing the income differential between the Congo and the United States purely as the result of a historical accident seems unsatisfactory. Instead, the literature has focused on the more modest goal of showing that the interaction of non-convexities and historical accidents can have an effect on the sectoral composition of production or on its spatial distribution within a given country.

[Juhász \(2014\)](#) uses data on 19th-century France to show a causal effect of temporary trade protection during the Napoleonic wars on the long-run location of the cotton-spinning industry in France. She argues that the continental blockade that prohibited direct shipping between Britain and France during the war differentially affected the costs of shipping goods from Britain to different regions within France. In particular, regions in Northern France that historically had relatively low costs of trading with Britain were temporarily more protected from British imports, because these imports now had to be shipped through Spain. Under the identifying assumption that more and less protected regions would have developed similarly absent the effect of the continental blockade on trade costs, she shows that relatively more protected regions significantly increased their capacity in mechanized cotton spinning (a new technology at the time). She then shows that this change in the regional distribution of the cotton-spinning industry within France persisted 30 years after the end of the blockade. Her results are thus consistent with non-convexities in the adoption of new technologies, and the view that infant industry protection can have a lasting effect on the location of production within and possibly also

across countries.⁵⁰

A large body of work in urban economics has used natural experiments to examine this question at the scale of cities. Although the focus of this literature is not explicitly on macroeconomic variables, we summarize it briefly, keeping in mind that factors of production, in particular labor, are much less mobile at the country or region level, such that it is unclear how results would change at larger levels of aggregation. In an influential study, [Davis and Weinstein \(2002, 2008\)](#) use a difference-in-differences approach to show that the bombing of Japanese cities during World War II had no long-run effect on the relative size of cities and even their industrial composition, suggesting that fundamentals rather than chance govern the long-run spatial distributions of these variables. [Miguel and Roland \(2011\)](#) show similar results using an instrumentation strategy for the US bombing of Vietnam.

[Bleakly and Lin \(2012\)](#) show contrasting evidence that a temporary locational advantage can have long-lasting effects on population density. They argue that many cities in North America formed in places where natural obstacles, such as waterfalls, blocked continued water transport and thus required overland hauling. These places (“portage sites”) attracted transportation services and commerce such that large settlements and cities would often form. However, this natural advantage is no longer relevant today, because trains, trucks, and airplanes have supplanted ships as the primary transportation technology, and locks and canals now make hauling cargo from one ship to another unnecessary. In this sense, technological progress generated a temporary positive shock to locational fundamentals of portage sites that plausibly no longer exists today. Nevertheless, [Bleakly and Lin](#) show not only that portage sites are significantly associated with higher population density today (including large cities such as Washington DC and Philadelphia), but also that no evidence exists of a relative decline in these areas, because their natural advantage as portage sites dissipated. Their findings are thus consistent with the existence of multiple equilibria in the location of cities and towns.

A series of papers in this literature also uses the division and reunification of Germany as a natural experiment. Most directly focused on testing for multiple equilibria is [Redding, Sturm, and Wolf \(2011\)](#). Using a simple difference-in-differences approach on a panel of passenger-traffic data for German airports, they show that the division of Germany led to a significant increase in the growth rate of passenger traffic in Frankfurt (previously a minor destination) and to a simultaneous shrinking of passenger traffic at

⁵⁰See [Kline and Moretti \(2014\)](#) for similar evidence from the Tennessee Valley Authority, where subsidies appear to have permanently increased the level of manufacturing employment.

Berlin airport (the previous main hub). However, after reunification, this trend did not reverse, leaving Germany's main airport hub in Frankfurt. Using various methods the authors argue that this apparently permanent shift cannot be explained by fundamentals, and instead is evidence of multiple equilibria in the location of a country's main airport hub.

Two closely related papers document evidence that is consistent with a specific source of multiple equilibria – the agglomeration externalities that operate in new economic geography models. The basic idea in this literature is that consumers like locating close to firms, and that firms like being close to other firms for various reasons, most commonly to economize on commuting and trade costs (Krugman, 1991). If these agglomeration forces are strong enough relative to dispersion forces (e.g., the costs of congestion), multiple equilibria may arise. One simple prediction of these models is that market access is an important driver of economic development. Redding and Sturm (2008) test this hypothesis using German division and reunification. Using a difference-in-differences approach, they find that over the 40-year period of German division, the population of West German cities close to the inner-German border declined significantly relative to other West German cities, a finding that is consistent with the predictions of the model. Their main identifying assumption is that, absent the effect of the inner-German border on market access, cities closer to and farther away from the border would have developed similarly. In Ahlfeldt, Redding, Sturm, and Wolf (2015), the authors take the analysis one step further and use the division and reunification of Berlin as a natural experiment, allowing them to quantitatively estimate agglomeration and dispersion forces at the city level. This paper is methodologically distinct from all the other papers covered in this chapter in that it focuses explicitly on structural, rather than reduced-form estimation. The authors develop a quantitative model of city structure that features multiple agglomeration and dispersion forces. They are able to separately identify these forces, because of the variation in the surrounding economic activity of city blocks that results from the division and reunification of the city. Interestingly, the identifying assumption for this structural model is just a sharper version of that used in the previous paper: that Berlin's division and subsequent reunification affects the systematic change in the pattern of economic activity across city blocks only through its effect on commuting costs and changes in access to production and residential externalities.

Taken together, these studies seem to suggest that historical accidents can have an effect on the spatial distribution of production within a given country. However, it is important to note that, conceptionally, none of these studies can distinguish multiple

equilibria from long-term persistence. Moreover, none of them speak to the existence of multiple equilibria at the country level.

5 Critical Assessment and Outlook

In this chapter, we describe the use of natural experiments in macroeconomics for three distinct purposes: to verify underlying model premises, to quantify policy parameters, and to identify causal mechanisms that are absent from conventional models. We do this by covering the use of natural experiments in the literatures on the Permanent Income Hypothesis, on fiscal multipliers, and on the fundamental causes of growth.

An easy test of the assumption of forward-looking behavior of the Permanent Income Hypothesis can be carried out if one can identify instances when households receive payments resulting from a preannounced change in income. A series of natural experiments that identify such preannounced changes finds that agents adjust consumption at receipt, which is in contrast to the assumption of forward-looking rational behavior. Some of these experiments suggest binding liquidity constraints can partly, but not entirely, explain this finding. They tend to find that more liquidity constrained households react more strongly at receipt, but even unconstrained households appear to adjust their consumption upon receiving a preannounced payment. The pattern of results across different types of natural experiments suggests a degree of near-rational behavior: in response to large income changes, households do appear to adhere to the Permanent Income Hypothesis, and only when faced with small income changes does the evidence not line up with the model predictions. Modelling near-rationality explicitly and embedding a natural experiment directly into such a model would be useful to generate further insights. Moreover, analyzing new measures of liquidity constraints that incorporate wealthy hand-to-mouth consumers could possibly help reconcile some of the conflicting evidence on the importance of liquidity constraints. A clear disadvantage of natural experiments also becomes apparent in this application: analyzing preannounced income *decreases* would directly rule out the possibility that binding liquidity constraints drive the results, and analyzing *large* preannounced income changes would rule out near-rational behavior. Whereas a researcher running field or laboratory experiments would thus make sure any given experiment covers both features, natural episodes encompassing these two features are rare.⁵¹

⁵¹Replicating large income changes in a designed experiment would, however, be costly and thus also difficult.

The literature on the fiscal multiplier faces the challenge of identifying fiscal policies that are orthogonal to current economic conditions. Two approaches involving natural experiments can help: the first uses increases in military spending caused by geopolitical events as exogenous shocks to national fiscal policies, and the second exploits cross-regional variation at the sub-national level driven by historical episodes to establish exogeneity. Whereas the first approach typically estimates aggregate fiscal multipliers smaller than 1, the second approach consistently returns estimates of the local fiscal multiplier between 1.5 and 2. However, these aggregate and regional multipliers are different concepts and cannot be directly compared. A model is needed to gain further insights into their relationship.⁵²

Identifying the fundamental reasons some countries are rich while others are poor is a major challenge in economics. The main empirical difficulty is that most of the likely drivers of growth, such as the accumulation of physical and human capital, the quality of institutions, the level of trust, and social structure, depend in complicated ways on the level of income and on each other. In the last two decades, natural experiments have become the most widely used tool to try to resolve this question by identifying the causal determinants of economic growth. Using such large-scale natural experiments as the age of European colonization or German reunification, the literature has produced convincing evidence that the major fundamental causes of growth are outside traditional models of economic growth. Instead, cross-country variation in political and economic institutions, cultural traits, and social structure appear to explain the majority of cross-country and often even the majority of within-country variation. The literature has been less successful at distinguishing these causal channels from each other. In particular, much of the work causally linking institutions and trust to growth does not convincingly differentiate between these two channels. In addition, much work remains to be done in trying to understand the dynamics and the interaction of these fundamental causes over time. The study of social structure, meaning social ties and networks as well as the social stratification of society, as a fundamental determinant of economic growth is a new, promising field that has arisen in this context, allowing a deeper understanding of the direct effect of social structure on growth and its effect on the evolution of institutions. A common thread across the different studies identifying the fundamental causes of economic growth is that they focus almost exclusively on the reduced-form relationship between the fundamental causes and economic outcomes, but make no use of the structure of existing

⁵²See the discussion of [Nakamura and Steinsson \(2014\)](#) in Section 3.

models of economic growth. Bridging the gap between these two literatures may prove a promising avenue for future research.

How researchers use natural experiments in practice differs between the three literatures we have covered. The literature on the Permanent Income Hypothesis employs natural experiments directly in identifying pre-announced income changes, whereas the other two strands of the literature mainly rely on microeconomic techniques such as instrumental variables and regression discontinuity designs. Some studies that do not rely on specific historical events for identification also use these techniques. As a result, deciding what can and cannot be considered a natural experiment in these latter two strands of literature involves some judgment.

The fundamental challenge, however, is the same in all three literatures, namely, to argue the historical episode in question provides the quasi-random variation that is necessary to identify causal effects. Although creating a comprehensive checklist that all analyses involving natural experiments can use is impossible, we list below some common features that distinguish successful papers that rely on the use of natural experiments:

Identifying assumption The identifying assumption underlying the study should be clearly stated. What does the reader need to believe to causally interpret the effect? What aspect of the natural experiment does the study exploit and how does it translate into exogenous variation? Why can reverse causality be ruled out? Can we think of any omitted factors that jointly affect the natural experiment and the outcome variable of interest?

Supporting evidence Researchers should provide as much evidence as possible to corroborate the identifying assumption. Such evidence may include showing treatment and control groups do not differ on observables prior to treatment, an analysis of pre- and post-treatment trends, the timing of the effect, and the use of multiple instruments. At a minimum, a careful description of the specific historical episode is needed that makes clear the origin of the quasi-random variation and allows the reader to gauge the potential scope for endogeneity and other confounding factors. Thus, detailed institutional knowledge on the episode surrounding the experiment is required. Researchers should carefully consider and explicitly discuss which omitted factors one might reasonably worry about.

Additional methods to support a causal interpretation If possible, researchers should use additional analyses to establish that the treatment and not other factors cor-

related with treatment drives the outcome variable. Specifically, placebo exercises that analyze placebo treatments can provide credibility to the claim of causality. Moreover, showing robustness to the use of different control groups if more than one is possible, and using matching methods to guarantee the control group lines up well with the treatment group in terms of observable characteristics can help assuage concerns about randomness.

Quantitative implications Arguing convincingly in favor of a causal interpretation is usually impossible without explicitly stating the quantitative implications of the estimates. Is the size of the effect reasonable? How does it compare to the influence of other factors that are known to affect the outcome variable?

A common promising avenue for future research identified in all three strands of literature is to move beyond reduced form approaches and incorporate natural experiments into structural models. Models will be useful to identify specific mechanisms at work and channels through which effects transmit, and to argue more forcefully about the plausibility of the quantitative sizes of effects. Natural experiments could then potentially prove useful in identifying a wider variety of structural model parameters.

Our discussion of the use of natural experiments in macroeconomics has also shown the “natural” limits of this approach. Because the researcher does not create the historical episode constituting the experiment, it is often not ideally suited to analyze the question at hand. As a result, different natural experiments often produce different answers to the same question. In addition, many of the more tightly identified natural experiments leave open the question of external validity. Finally, because natural experiments rely on history for identification, they are naturally more useful for understanding the past rather than the future. Although this is not problematic in many contexts, it does preclude natural experiments from addressing the effects of unprecedented events, such as climate change.

Nevertheless, we believe that two decades of research using natural experiments have produced a range of substantial insights into the functioning of the economy, and have made major contributions to the field of macroeconomics.

References

- Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program. *Journal of the American Statistical Association* 105(490), 493–505.
- Abadie, A. and J. Gardeazabal (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review* 93(1), 113–132.
- Abdallah, C. S. and W. D. Lastrapes (2012). Home Equity Lending and Retail Spending: Evidence from a Natural Experiment in Texas. *American Economic Journal: Macroeconomics* 4(4), 94–125.
- Acconcia, A., G. Corsetti, and S. Simonelli (2014). Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-experiment. *American Economic Review* 104(7), 2185–2209.
- Acemoglu, D. (2009). *Introduction to Modern Economic Growth*. Princeton University Press.
- Acemoglu, D., D. Cantoni, S. Johnson, and J. A. Robinson (2011). The Consequences of Radical Reform: The French Revolution. *American Economic Review* 101(7), 3286–3307.
- Acemoglu, D., F. A. Gallego, and J. A. Robinson (2014). Institutions, Human Capital, and Development. *Annual Review of Economics* 6(875-912).
- Acemoglu, D., T. A. Hassan, and J. A. Robinson (2011). Social Structure and Development: A Legacy of the Holocaust in Russia. *Quarterly Journal of Economics* 126(2), 895–946.
- Acemoglu, D., T. A. Hassan, and A. Tahoun (2014). The Power of the Street: Evidence from Egypt’s Arab Spring. *mimeo University of Chicago*.
- Acemoglu, D., S. Johnson, J. Robinson, and Y. Thaicharoen (2003). Institutional causes, macroeconomic symptoms: volatility, crises and growth. *Journal of Monetary Economics* 50(1), 49–123.

- Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review* 91, 1369–1401.
- Acemoglu, D., S. Johnson, and J. A. Robinson (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *Quarterly Journal of Economics* 117, 1231–1294.
- Acemoglu, D., S. Johnson, and J. A. Robinson (2005). The Rise of Europe: Atlantic Trade, Institutional Change and Economic Growth. *American Economic Review* 95, 546–579.
- Acemoglu, D., S. Johnson, and J. A. Robinson (2012). The Colonial Origins of Comparative Development: An Empirical Investigation: Reply. *American Economic Review* 102(6), 3077–3110.
- Acemoglu, D., S. Johnson, J. A. Robinson, and P. Yared (2005). From Education to Democracy. *American Economic Review Papers and Proceedings* 95(2), 44–49.
- Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (2014). Democracy does cause growth. Technical report, National Bureau of Economic Research.
- Acemoglu, D. and J. A. Robinson (2005). *Economic origins of dictatorship and democracy*. Cambridge University Press.
- Agarwal, S., C. Liu, and N. S. Souleles (2007). The Reaction of Consumer Spending and Debt to Tax Rebates-Evidence from Consumer Credit Data. *Journal of Political Economy* 115(6), 986–1019.
- Agarwal, S., N. Marwell, and L. McGranahan (2013). Consumption Responses to Temporary Tax Incentives: Evidence from State Sales Holidays. Working Paper.
- Agarwal, S. and W. Qian (2014a). Consumption and Debt Response to Unanticipated Income Shocks: Evidence from a Natural Experiment in Singapore. *American Economic Review* 104(12), 4205–30.
- Agarwal, S. and W. Qian (2014b). Access to Home Equity and Consumption: Evidence from a Policy Experiment. Working Paper.

- Aghion, P., Y. Algan, P. Cahuc, and A. Shleifer (2010). Regulation and Distrust. *Quarterly Journal of Economics* 125(3), 1015–1049.
- Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015). The Economics of Density: Evidence from the Berlin Wall. *Econometrica* (forthcoming).
- Akerlof, G. and J. Yellen (1985). A Near-Rational Model of the Business Cycle with Wage and Price Inertia. *Quarterly Journal of Economics*, 823–838.
- Albouy, D. Y. (2012). The Colonial Origins of Comparative Development: An Empirical Investigation: Comment. *American Economic Review* 102(6), 3059–3076.
- Alesina, A. and N. Fuchs-Schündeln (2007). Good Bye Lenin (or Not?): The Effect of Communism on People. *American Economic Review* 97(4), 1507–1528.
- Alesina, A. and P. Giuliano (2015). Culture and institutions. *Journal of Economic Literature* (forthcoming).
- Algan, Y. and P. Cahuc (2010). Inherited Trust and Growth. *American Economic Review* 100(5), 2060–2092.
- Algan, Y. and P. Cahuc (2013). Trust and Growth. *Annual Review of Economics* 5(1), 521–549.
- Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy* 113(1), 151–184.
- Anderson, C. J. (2003). The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psychological Bulletin* 129, 139–167.
- Arrow, K. J. (1972). Gifts and Exchanges. *Philosophy & Public Affairs* 1(4), pp. 343–362.
- Auerbach, A. J. and Y. Gorodnichenko (2012, May). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy* 4(2), 1–27.
- Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2013). The Diffusion of Microfinance. *Science* 341(6144).

- Banerjee, A. and L. Iyer (2005). History, Institutions and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. *American Economic Review* 95(4), 1190–1213.
- Banfield, E. C. (1967). *The moral basis of a backward society*. Free Press.
- Barro, R. J. (1981). Output Effects of Government Purchases. *Journal of Political Economy* 89(6), 1086–1121.
- Barro, R. J. (1999). Determinants of democracy. *Journal of Political Economy* 107(S6), S158–S183.
- Barro, R. J. and C. J. Redlick (2011). Macroeconomic Effects From Government Purchases and Taxes. *The Quarterly Journal of Economics* 126(1), 51–102.
- Beaman, L. (2012). Social Networks and the Dynamics of Labor Market Outcomes: Evidence From Refugees Resettled in the U.S. *The Review of Economic Studies* 79(1), 128–161.
- Becker, S. O., K. Boeckh, C. Hainz, and L. Woessmann (2015). The Empire is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy. *The Economic Journal* (forthcoming).
- Becker, S. O., P. H. Egger, and M. von Ehrlich (2010). Going NUTS: The effect of EU Structural Funds on regional performance. *Journal of Public Economics* 94(9-10), 578 – 590.
- Becker, S. O., P. H. Egger, and M. von Ehrlich (2013). Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects. *American Economic Journal: Economic Policy* 5(4), 29–77.
- Bellows, J. and E. Miguel (2009). War and local collective action in Sierra Leone. *Journal of Public Economics* 93(11-12), 1144 – 1157.
- Bertrand, M., E. Luttmer, and S. Mullainathan (2000). Network Effects and Welfare Cultures. *The Quarterly Journal of Economics* 115(3), 1019–1055.
- Bertrand, M. and A. Morse (2009). What Do High-Interest Borrowers Do with Their Tax Rebate? *American Economic Review* 99(2), 418–423.

- Blanchard, O. and R. Perotti (2002). An Empirical Characterization Of The Dynamic Effects Of Changes In Government Spending And Taxes On Output. *The Quarterly Journal of Economics* 117(4), 1329–1368.
- Bleakely, H. and J. Lin (2012). Portage and Path Dependence. *Quarterly Journal of Economics* 127, 587–644.
- Bloom, N. (2009). The Impact of Uncertainty Shocks. *Econometrica* 77(3), 623–685.
- Bodkin, R. (1959). Windfall Income and Consumption. *The American Economic Review* 49(4), pp. 602–614.
- Broda, C. and J. A. Parker (2014). The Economic Stimulus Payments of 2008 and the Aggregate Demand for Consumption. *Journal of Monetary Economics* 68, Supplement(0), 20–36.
- Browning, M. and M. D. Collado (2001). The Response of Expenditures to Anticipated Income Changes: Panel Data Estimates. *American Economic Review* 91(3), 681–692.
- Brückner, M. and A. Tuladhar (2014). Local Government Spending Multipliers and Financial Distress: Evidence from Japanese Prefectures. *The Economic Journal* 124(581), 1279–1316.
- Brueckner, M. and A. Ciccone (2011). Rain and the democratic window of opportunity. *Econometrica* 79(3), 923–947.
- Brueckner, M., A. Ciccone, and A. Tesei (2012). Oil Price Shocks, Income, and Democracy. *Review of Economics and Statistics* 94(2), 389–3.
- Brueckner, M. and M. Gradstein (2013). Effects of transitory shocks to aggregate output on consumption in poor countries. *Journal of International Economics* 91, 343–357.
- Burchardi, K. B., T. Chaney, and T. A. Hassan (2015). Migrants, Trade, and Investments. *Working Paper*.
- Burchardi, K. B. and T. A. Hassan (2013). The Economic Impact of Social Ties: Evidence from German Reunification. *Quarterly Journal of Economics* 128(3), 1219–1271.
- Burke, P. J. and A. Leigh (2010). Do Output Contractions Trigger Democratic Change? *American Economic Journal: Macroeconomics* 2(4), 124–57.

- Bursztyn, L. and D. Cantoni (2015). A Tear in the Iron Curtain: The Impact of Western Television on Consumption Behavior. *Review of Economics and Statistics* (forthcoming).
- Burt, R. (1992). *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge MA.
- Carroll, C. D. and L. H. Summers (1991, January). *Consumption Growth Parallels Income Growth: Some New Evidence*, pp. 305–348. University of Chicago Press.
- Caselli, F. and A. Tesei (2015). Resource windfalls, political regimes, and political stability. Technical report. Forthcoming, Review of Economics and Statistics.
- Chaney, E. (2013). Revolt on the Nile: Economic Shocks, Religion and Political Power. *Econometrica* 81(5), 2033–2053.
- Chaney, T. (2014). *Oxford Handbook of the Economics of Networks*, Chapter Networks in International Trade.
- Chetty, R. (2012). Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply. *Econometrica* 80(3), 969–1018.
- Chodorow-Reich, G., L. Feiveson, Z. Liscow, and W. G. Woolston (2012). Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act. *American Economic Journal: Economic Policy* 4(3), 118–45.
- Clemens, J. and S. Miran (2012). Fiscal Policy Multipliers on Subnational Government Spending. *American Economic Journal: Economic Policy* 4(2), 46–68.
- Cochrane, J. (1989). The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives. *The American Economic Review* 79(3), 319–37.
- Cohen, L., J. Coval, and C. Malloy (2011). Do Powerful Politicians Cause Corporate Downsizing? *Journal of Political Economy* 119(6), 1015 – 1060.
- Cohen, L., A. Frazzini, and C. Malloy (2008). The Small World of Investing: Board Connections and Mutual Fund Returns. *Journal of Political Economy* 116(5), 951–979.

- Cohen, L., U. Gurun, and C. Malloy (2014). Resident Networks and Firm Trade. Working Paper, October 2014.
- Coleman, J. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology* 94, 95–120.
- Combes, P.-P., M. Lafourcade, and T. Mayer (2005). The trade-creating effects of business and social networks: evidence from France. *Journal of International Economics* 66, 1–29.
- Conley, T. and C. Udry (2010). Learning About a New Technology: Pineapple in Ghana. *The American Economic Review* 100(1), 35–69.
- Corbi, R., E. Papaioannou, and P. Surico (2014). Federal Transfer Multipliers. Quasi-Experimental Evidence from Brazil. Working Paper 20751, National Bureau of Economic Research.
- Coulibaly, B. and G. Li (2006). Do Homeowners Increase Consumption after the Last Mortgage Payment? An Alternative Test of the Permanent Income Hypothesis. *Review of Economics and Statistics* 88(1), 10–19.
- Davis, D. R. and D. E. Weinstein (2002). Bones, Bombs, and Breakpoints: The Geography of Economic Activity. *American Economic Review* 92(5), 1269–1289.
- Davis, D. R. and D. E. Weinstein (2008). A Search for Multiple Equilibria in Urban Industrial Structure. *Journal of Regional Science* 48(1), 29–65.
- DeFusco, A. A. (2014). Homeowner Borrowing and Housing Collateral: New Evidence from Expiring Price Controls. Mimeo, University of Pennsylvania.
- Dell, M. (2010). The Persistent Effects of Peru’s Mining Mita. *Econometrica* 78(6), 1863–1903.
- Dippel, C. (2014). Forced Coexistence and Economic Development: Evidence from Native American Reservations. *Econometrica* 82(6), 2131–2165.
- Duffy, J. (2008). Macroeconomics: a survey of laboratory research. *Handbook of Experimental Economics* 2.
- Durante, R. (2010). Risk, Cooperation and the Economic Origins of Social Trust: An Empirical Investigation. Mimeo, Science Po.

- Easterly, W. and R. Levine (2003). Tropics, Germs and Crops: How Endowments Influence Economic Development. *Journal of Monetary Economics* 50, 3–39.
- Easterly, W. and R. Levine (2012). The European Origins of Economic Development. NBER Working Paper 18162.
- Edelberg, W., M. Eichenbaum, and J. D. Fisher (1999). Understanding the Effects of a Shock to Government Purchases. *Review of Economic Dynamics* 2(1), 166–206.
- Evans, W. N. and T. J. Moore (2011). The short-term mortality consequences of income receipt. *Journal of Public Economics* 95(11), 1410–1424.
- Feyrer, J. D. and B. Sacerdote (2012). Did the Stimulus Stimulate? Effects of the American Recovery and Reinvestment Act. Mimeo, Dartmouth University.
- Fishback, P. and J. A. Cullen (2013). Second World War spending and local economic activity in US counties, 1939-58. *Economic History Review* 66(4), 975–992.
- Fishback, P. V. and V. Kachanovskaya (2015). In Search of the Multiplier for Federal Spending in the States During the Great Depression. *The Journal of Economic History* 75(1), 125–162.
- Fisman, R. (2001). Estimating the Value of Political Connections. *American Economic Review* 91(4), 1095–1102.
- Franck, R. (2015). The political consequences of income shocks: Explaining the consolidation of democracy in france. *Review of Economics and Statistics* (forthcoming).
- Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton University Press.
- Fuchs-Schündeln, N. (2008). The Response of Household Saving to the Large Shock of German Reunification. *American Economic Review* 98(5), 1798–1828.
- Fuchs-Schündeln, N. and M. Schündeln (2005). Precautionary Savings and Self-Selection: Evidence from the German Reunification “Experiment”. *The Quarterly Journal of Economics* 120, 1085–1120.
- Fuchs-Schündeln, N. and M. Schündeln (2015). On the Endogeneity of Political Preferences: Evidence from Individual Experience with Democracy. *Science* 347(6226), 1145–1148.

- Garmendia, A., C. Llano, A. Minondo, and F. Requena (2012). Networks and the disappearance of the intranational home bias. *Economics Letters* 116, 178–182.
- Gebhardt, G. (2013). Does Relationship Specific Investment Depend on Asset Ownership? Evidence from a Natural Experiment in the Housing Market. *Journal of the European Economic Association* 11(1), 201–227.
- Gelman, M., S. Kariv, M. D. Shapiro, D. Silverman, and S. Tadelis (2014). Harnessing naturally occurring data to measure the response of spending to income. *Science* 345(6193), 212–215.
- Glaeser, E., D. Laibson, and B. Sacerdote (2002). An Economic Approach to Social Capital*. *The Economic Journal* 112(483), F437–F458.
- Glaeser, E. L., R. La Porta, F. Lopez-De-Silanes, and A. Shleifer (2004). Do Institutions Cause Growth? *Journal of Economic Growth* 9, 271–303.
- Glaeser, E. L., G. Ponzetto, and A. Shleifer (2007). Why does democracy need education? *Journal of Economic Growth* 12(2), 77–99.
- Gorodnichenko, Y. and G. Roland (2010). Culture, institutions and the wealth of nations. Technical report, National Bureau of Economic Research.
- Gould, D. M. (1994). Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows. *Review of Economics and Statistics* 76(2), 302–316.
- Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology* 91 (3), 481–510.
- Granovetter, M. (2005). The Impact of Social Structure on Economic Outcomes. *The Journal of Economic Perspectives* 19 (1), 33–50.
- Greif, A. (1993a). Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders’ Coalition. *The American Economic Review* 83(3), 525–548.
- Greif, A. (1993b). Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders’ Coalition. *American Economic Review* 83(3), 525–548.
- Gross, T., M. J. Notowidigdo, and J. Wang (2014). Liquidity Constraints and Consumer Bankruptcy: Evidence from Tax Rebates. *Review of Economics and Statistics* 96(3), 431–443.

- Gross, T. and J. Tobacman (2014). Dangerous Liquidity and the Demand for Health Care: Evidence from the 2008 Stimulus Payments. *Journal of Human Resources* 49(2), 424–445.
- Guiso, L., P. Sapienza, and L. Zingales (2004). The Role of Social Capital in Financial Development. *American Economic Review* 94(3), 526–556.
- Guiso, L., P. Sapienza, and L. Zingales (2006). Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives* 20(2), 23–48.
- Guiso, L., P. Sapienza, and L. Zingales (2008a). Long-Term Persistence. NBER Working Paper No. 14278.
- Guiso, L., P. Sapienza, and L. Zingales (2008b). Social Capital as Good Culture. *Journal of the European Economic Association* 6(2-3), 295–320.
- Guiso, L., P. Sapienza, and L. Zingales (2012). *The Handbook of Social Economics*, Chapter Civic Capital as the Missing Link, pp. 417–480. Elsevier.
- Hall, R. E. (1986, May). *The Role of Consumption in Economic Fluctuations*, pp. 237–266. NBER Chapters. University of Chicago Press.
- Hall, R. E. (2009). By How Much Does GDP Rise If the Government Buys More Output? *Brookings Papers on Economic Activity* 40(2), 183–249.
- Hall, R. E. and C. I. Jones (1999). Why Do Some Countries Produce So Much More Output Per Worker Than Others? *The Quarterly Journal of Economics* 114(1), 83–116.
- Hassan, T. and T. Mertens (2014). The Social Cost of Near-Rational Investment. *NBER Working Paper*.
- Hausman, J. K. (2015). Fiscal Policy and Economic Recovery: The Case of the 1936 Veterans’ Bonus. *The American Economic Review* (forthcoming).
- Hochberg, Y., A. Ljungqvist, and Y. Lu (2007). Whom You Know Matters: Venture Capital Networks and Investment Performance. *The Journal of Finance* 62, 251–301.
- Hornbeck, R. and S. Naidu (2014). When the Levee Breaks: Black Migration and Economic Development in the American South. *American Economic Review* 104(3), 963–90.

- Hsieh, C.-T. (2003). Do Consumers React to Anticipated Income Changes? Evidence from the Alaska Permanent Fund. *American Economic Review* 93(1), 397–405.
- Huntington, S. P. (1991). Democracy’s third wave. *Journal of Democracy* 2(2), 12–34.
- Imbens, G. W., D. B. Rubin, and B. I. Sacerdote (2001). Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players. *American Economic Review* 91(4), 778–794.
- Iyer, L. (2010). Direct Versus Indirect Colonial Rule in India: Long-Term Consequences. *Review of Economics and Statistics* 92(4), 693–713.
- Jancec, M. (2012). Do Less Stable Borders Lead to Lower Levels of Political Trust? Empirical Evidence from Eastern Europe. Mimeo, University of Maryland at College Park.
- Jappelli, T. and L. Pistaferri (2010). The Consumption Response to Income Changes. *Annual Review of Economics* 2, 479–506.
- Johnson, D. S., J. A. Parker, and N. S. Souleles (2006). Household Expenditure and the Income Tax Rebates of 2001. *American Economic Review* 96(5), 1589–1610.
- Jones, E. (2003). *The European miracle: environments, economies and geopolitics in the history of Europe and Asia*. Cambridge University Press.
- Juhasz, R. (2014). Temporary Protection and Technology Adoption: Evidence from the Napoleonic Blockade. Job Market Paper.
- Kaplan, G. and G. I. Violante (2014a). A Model of the Consumption Response to Fiscal Stimulus Payments. *Econometrica* 82(4), 1199–1239.
- Kaplan, G. and G. I. Violante (2014b). A Tale of Two Stimulus Payments: 2001 vs. 2008. *American Economic Review* 104(5), 116–121.
- Kline, P. and E. Moretti (2014). Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. *The Quarterly Journal of Economics* 129(1), 275–331.
- Knack, S. and P. Keefer (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *Quarterly Journal of Economics* 112(4), 1251–1288.

- Kraay, A. (2012). How large is the Government Spending Multiplier? Evidence from World Bank Lending. *The Quarterly Journal of Economics* 127(2), 829–887.
- Kraay, A. (2014). Government Spending Multipliers in Developing Countries: Evidence from Lending by Official Creditors. *American Economic Journal: Macroeconomics* 6(4), 170–208.
- Kreinin, M. E. (1961). Windfall Income and Consumption: Additional Evidence. *The American Economic Review* 51(3), pp. 388–390.
- Krugman, P. (1991). Increasing Returns and Economic Geography. *The Journal of Political Economy* 99(3), 483–499.
- Kueng, L. (2015). Explaining Consumption Excess Sensitivity with Near-Rationality: Evidence from Large Predetermined Payments. *mimeo*.
- Kuhn, P., P. Kooreman, A. Soetevent, and A. Kapteyn (2011). The Effects of Lottery Prizes on Winners and Their Neighbors: Evidence from the Dutch Postcode Lottery. *American Economic Review* 101(5), 2226–47.
- Kuhnen, C. (2009). Business Networks, Corporate Governance, and Contracting in the Mutual Fund Industry. *The Journal of Finance* 64(5), 2185–2220.
- La Porta, R., F. Lopez-de-Silanes, and A. Shleifer (2008). The Economic Consequences of Legal Origins. *Journal of Economic Literature* 46(2), 285–332.
- La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny (1997). Legal Determinants of External Finance. *Journal of Finance* 52, 1131–1150.
- La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. Vishny (1998). Law and Finance. *Journal of Political Economy* 106, 1113–1155.
- Landsberger, M. (1966). Windfall Income and Consumption: Comment. *The American Economic Review* 56(3), pp. 534–540.
- Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *American Political Science Review* 53(01), 69–105.
- Loury, G. C. (1977). *Women, Minorities and Employment Discrimination*, Chapter A Dynamic Theory of Racial Income Differences, pp. 153–188. Lexington Books.

- Ludvigson, S. C. and A. Michaelides (2001). Does Buffer-Stock Saving Explain the Smoothness and Excess Sensitivity of Consumption? *American Economic Review* 91(3), 631–647.
- Madestam, A., D. Shoag, S. Veuger, and D. Yanagizawa-Drott (2013). Do Political Protests Matter? Evidence from the Tea Party Movement. *The Quarterly Journal of Economics* 128(4), 1633–1685.
- Mankiw, N. G. (1985). Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly. *The Quarterly Journal of Economics* 100(2), 529–537.
- Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* 60, 531–542.
- Mastrobuoni, G. and M. Weinberg (2009). Heterogeneity in Intra-monthly Consumption Patterns, Self-Control, and Savings at Retirement. *American Economic Journal: Economic Policy* 1(2), 163–89.
- Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics* 110(3), 681–712.
- Mian, A. and A. Sufi (2012). The Effects of Fiscal Stimulus: Evidence from the 2009 Cash for Clunkers Program. *The Quarterly Journal of Economics* 127(3), 1107–1142.
- Michalopoulos, S. and E. Papaioannou (2011). The Long-Run Effects of the Scramble for Africa. NBER Working Paper 17620.
- Michalopoulos, S. and E. Papaioannou (2014). National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics* 129(1), 151–213.
- Miguel, E. and G. Roland (2011). The Long Run Impact of Bombing Vietnam. *Journal of Development Economics* 96(1), 1–15.
- Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic Shocks and Civil Conflict: An Instrumental Variables Approach. *Journal of Political Economy* 112(4), pp. 725–753.
- Misra, K. and P. Surico (2014). Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs. *American Economic Journal: Macroeconomics* 6(4), 84–106.

- Mountford, A. and H. Uhlig (2009). What are the effects of fiscal policy shocks? *Journal of Applied Econometrics* 24(6), 960–992.
- Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market. *The Quarterly Journal of Economics* 118(2), 549–599.
- Murphy, K. M., A. Shleifer, and R. W. Vishny (1989). Industrialization and the Big Push. *Journal of Political Economy* 97, 1003–1026. Reprinted in Dilip Mookherjee and Debraj Ray, eds., *Readings in Theory of Economic Development*, Blackwell Publishing, 2001.
- Nakamura, E. and J. Steinsson (2014). Fiscal Stimulus in a Monetary Union: Evidence from US Regions. *American Economic Review* 104(3), 753–92.
- North, D. C. (1981). *Structure and change in economic history*. W.W. Norton & Company.
- North, D. C. and R. P. Thomas (1973). *The Rise of the Western World: a New Economic History*. Cambridge University Press.
- Nunn, N. (2013). Historical development. *Handbook of Economic Growth* 2, 347.
- Nunn, N. and L. Wantchekon (2011). The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review* 101(7), 3221–3252.
- Olsson, O. (2004). Unbundling Ex-Colonies: A Comment on Acemoglu, Johnson, and Robinson, 2001. Mimeo, Goteborg University.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
- Owyang, M. T., V. A. Ramey, and S. Zubairy (2013). Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data. *American Economic Review* 103(3), 129–34.
- Parker, J. A. (2014). Why Don't Households Smooth Consumption? Evidence from a 25 Million Dollar Experiment. *mimeo*.
- Parker, J. A., N. S. Souleles, D. S. Johnson, and R. McClelland (2013). Consumer Spending and the Economic Stimulus Payments of 2008. *American Economic Review* 103(6), 2530–53.

- Parsons, C. and P.-L. Vezina (2014). Migrant Networks and Trade: The Vietnamese Boat People as a Natural Experiment. *mimeo, University of Oxford*.
- Paxson, C. H. (1993). Consumption and Income Seasonality in Thailand. *Journal of Political Economy* 101(1), pp. 39–72.
- Persson, T. and G. Tabellini (2009). Democratic Capital: The Nexus of Political and Economic Change. *American Economic Journal: Macroeconomics* 1(2), 88–126.
- Pinkovskiy, M. L. (2013). Economic discontinuities at borders: Evidence from satellite data on lights at night. Mimeo, Federal Reserve Bank of New York.
- Portes, R. and H. Rey (2005). The determinants of cross-border equity flows. *Journal of International Economics* 65(2), 269–296.
- Putnam, R., R. Leonardi, and R. Nanetti (1993). *Making Democracy Work*. New York: Simon & Schuster.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon and Schuster.
- Ramey, V. A. (2011). Identifying Government Spending Shocks: It’s all in the Timing. *The Quarterly Journal of Economics* 126(1), 1–50.
- Ramey, V. A. and M. D. Shapiro (1998, June). Costly capital reallocation and the effects of government spending. *Carnegie-Rochester Conference Series on Public Policy* 48(1), 145–194.
- Rauch, J. E. and V. Trindade (2002). Ethnic Chinese Networks in International Trade. *Review of Economics and Statistics* 84(1), 116–130.
- Redding, S. J., D. Sturm, and N. Wolf (2011). History and Industrial Location: Evidence from German Airports. *Review of Economics and Statistics* 93(3), 814–831.
- Redding, S. J. and D. M. Sturm (2008). The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review* 98(5), 1766–97.
- Reis, R. (2006). Inattentive consumers. *Journal of Monetary Economics* 53(8), 1761 – 1800.

- Rice, T. W. and J. L. Feldman (1997). Civic Culture and Democracy from Europe to America. *The Journal of Politics* 59(4), pp. 1143–1172.
- Roberts, B. E. (1990). A Dead Senator Tells No Lies: Seniority and the Distribution of Federal Benefits. *American Journal of Political Science* 34(1), 31–58.
- Rodrik, D., A. Subramanian, and F. Trebbi (2004). Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. *Journal of Economic Growth* 9(2), 131–165.
- Romer, C. D. and D. H. Romer (2010). The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *American Economic Review* 100(3), 763–801.
- Rosenbaum, P. R. and D. B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70(1), 41–55.
- Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics* 116 (2), 681–704.
- Sahm, C. R., M. D. Shapiro, and J. Slemrod (2010, August). *Household Response to the 2008 Tax Rebate: Survey Evidence and Aggregate Implications*, pp. 69–110. The University of Chicago Press.
- Sahm, C. R., M. D. Shapiro, and J. Slemrod (2012). Check in the Mail or More in the Paycheck: Does the Effectiveness of Fiscal Stimulus Depend on How It Is Delivered? *American Economic Journal: Economic Policy* 4(3), 216–50.
- Saxenian, A. (1999). *Silicon Valley's New Immigrant Entrepreneurs*. Public Policy Institute of California.
- Scholnick, B. (2013). Consumption Smoothing after the Final Mortgage Payment: Testing the Magnitude Hypothesis. *Review of Economics and Statistics* 95(4), 1444–1449.
- Scott R. Baker and Nicholas Bloom (2013). Does uncertainty reduce growth? Using disasters as natural experiments. Technical report, National Bureau of Economic Research.
- Serrato, J. C. S. and P. Wingender (2014). Estimating Local Fiscal Multipliers. Mimeo, Duke University.

- Shapiro, J. M. (2005). Is there a daily discount rate? Evidence from the food stamp nutrition cycle. *Journal of Public Economics* 89(2-3), 303–325.
- Shapiro, M. D. and J. Slemrod (2003). Consumer Response to Tax Rebates. *American Economic Review* 93(1), 381–396.
- Shoag, D. (2013). Using State Pension Shocks to Estimate Fiscal Multipliers since the Great Recession. *American Economic Review* 103(3), 121–24.
- Shoag, D. (2015). The Impact of Government Spending Shocks: Evidence on the Multiplier from State Pension Plan Returns. Mimeo, Harvard University.
- Shue, K. (2013). Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers. *Review of Financial Studies* 26(6), 1401–1442.
- Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics* 70(1), 65–94.
- Souleles, N. S. (1999). The Response of Household Consumption to Income Tax Refunds. *American Economic Review* 89(4), 947–958.
- Souleles, N. S. (2000). College tuition and household savings and consumption. *Journal of Public Economics* 77(2), 185 – 207.
- Stephens, Melvin, J. (2003). "3rd of tha Month": Do Social Security Recipients Smooth Consumption between Checks? *The American Economic Review* 93(1), pp. 406–422.
- Stephens, Melvin, J. (2006, 07). Paycheque Receipt and the Timing of Consumption. *Economic Journal* 116(513), 680–701.
- Stiglitz, J. E. (1990). Peer Monitoring and Credit Markets. *The World Bank Economic Review* 4(3), 351–366.
- Swan, T. W. (1956). Economic growth and capital accumulation. *Economic Record* 32(2), 334–361.
- Tabellini, G. (2010). Culture and institutions: economic development in the regions of Europe. *Journal of the European Economic Association* 8(4), 677–716.
- Trezzi, R. and F. Porcelli (2014). Reconstruction multipliers. Finance and Economics Discussion Series 2014-79, Board of Governors of the Federal Reserve System (U.S.).

Wilson, D. J. (2012). Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act. *American Economic Journal: Economic Policy* 4(3), 251–82.

Appendix Table 1: Overview of Permanent Income Hypothesis Papers Relying on Natural Experiments

Study	Experiment	Data Source	Sample	Main Dependent Variable	Frequency of Data	PIH	Quantitative Reaction at Implementation	Liquidity Constraint
Standard PIH studies								
Aaronson, Agarwal, and French (2012)	Minimum Wage Hikes; income increase	CEX interview survey (1982-2008); CPS (1980-2007) and SIPP (1983-2007) for data on income; proprietary dataset from national financial institution (1995-2008): credit card account and each credit card holder's auto, home equity, mortgage, and credit card balance	200,500 household-survey observations on spending, of which 11 percent derive some income from minimum wage work	spending on nondurables and durables, including separate sub-categories, and change in debt (total and sub-categories)	quarterly	reject	households where minimum wage labor is the source of at least 20 percent of household income spend 3.4 times the short-term increase in income, but mostly on cars	based on liquid assets; some evidence in favor
Agarwal, Liu, and Souleles (2007)	2001 Federal Income Tax Rebates; income increase	proprietary data set from a large financial institution that issues credit cards nationally (2000-2002): data on credit bureau reports, credit card spending, balance, debt	75,000 credit card accounts open as of June 2000, followed monthly for 24 months	credit card spending, balances, and debt, as well as credit limit	monthly	reject	spending increases by 40% of the average household rebate cumulatively in the 9 months after receipt for consumers whose most intensively used credit card account is in the sample	based on credit limit, utilization rate, and age; strong evidence in favor
Agarwal and Qian (2014)	2011 Singapore Growth Dividends; income increase	credit card, debit card, and bank checking account data from leading bank in Singapore (2010-2012)	random sample of all bank customers of leading Singaporean bank (180,000 observations)	spending, debt, credit card usage	monthly	cannot reject	increase in card spending by 8 cents per month for every dollar received, corresponding to a total increase of 80 cents in the 10 month period after announcement; monthly increase similar after announcement and implementation	based on liquid assets and credit card limit; liquidity constrained consumers react stronger, though also already at announcement, i.e. likely incompletely binding constraints
Broda and Parker (2014)	2008 Economic Stimulus Payments; income increase	Nielsen's Consumer Panel (2008): scanned purchases in grocery stores, drugstores and mass-merchandise sectors	data on weekly purchases of 28,937 households (1,131,208 observations)	weekly spending on household goods based on barcode scanners used by households	weekly	reject	three months cumulative increase in spending on Nielsen Consumer Panel goods amounts to 7 percent of ESP receipt	based on income and survey question regarding the availability of easily accessible funds; evidence in favor

Study	Experiment	Data Source	Sample	Main Dependent Variable	Frequency of Data	PIH	Quantitative Reaction at Implementation	Liquidity Constraint
Browning and Collado (2001)	Spanish Bonus Payment Scheme; income increase and decrease	Spanish Encuesta Continua de Presupuestos Familiares (1985-1995): data on earnings and spending	2,341 households (16,143 observations) of which about 80 percent are classified as bonus	spending on nondurables and durables, including separate sub-categories	weekly	cannot reject	spending patterns of bonus and non-bonus groups are indistinguishable	no test
Coulibaly and Li (2006)	Last mortgage payment; disposable income increase	CEX interview survey (1984-2000)	70,593 observations, including 286 with last mortgage payments	spending on nondurables and durables, plus subcategories	quarterly	cannot reject	homeowners do not increase consumption after the last mortgage payment	no test (none of respondents likely to be constrained)
Gelman, Kariv, Shapiro, Silverman, and Tadelis (2014)	regular paycheck or social security check arrival; income increase and decrease	Check, a financial aggregation and service application combining information from different financial accounts (2012-2013)	75,000 randomly sampled US Check users	total spending, nonrecurring spending, and fast food and coffee shop spending	daily	cannot reject	Nonrecurring spending and fast food and coffee shop spending show only mild comovement with regular payments	based on ratio of average daily balance on saving and checking account to average daily spending; strong evidence in favor
Hsieh (2003)	Alaska Permanent Fund Payments and income tax refunds; income increase and decrease	CEX interview survey (1980-1981, 1984-2001)	806 Alaskan households	spending on nondurables and durables, and different sub-categories	quarterly	cannot reject	a 10 percent increase in household income increases non-durable consumption insignificantly by only 0.002 percent	based on current income; no evidence in favor
Johnson, Parker, and Souleles (2006)	2001 Federal Income Tax Rebates; income increase	CEX interview survey (2000-2002) with added questions on tax rebates	13,066 observations on households who received the tax rebate	spending on nondurables, strictly nondurables, and food	quarterly	reject	households spend 20-40 percent of their rebates on non-durable consumption goods during the three-months period in which the rebates were received, and roughly two thirds during six-months period	based on age, income, and liquid assets; some evidence based on income and liquid assets, not age

Study	Experiment	Data Source	Sample	Main Dependent Variable	Frequency of Data	PIH	Quantitative Reaction at Implementation	Liquidity Constraint
Mastrobuoni and Weinberg (2009)	Social Security benefits payments; income increase and decrease	Continuing Survey of Food Intake by Individuals (1994-1996)	745 observations from households in which Social Security income makes up at least 80 percent of total income	caloric intake	daily	mixed	retirees with savings above \$5000 smooth caloric intake over the pay cycle, while those with less than \$5000 in savings have 24 percent lower caloric intake during the final few days of pay cycle than during first week	based on liquid assets; strong evidence in favor (see main result)
Misra and Surico (2014)	2001 Federal Income Tax Rebates and 2008 Economic Stimulus Payments; income increase	CEX interview survey (2000-2002 and 2007-2008) with added questions on tax rebate and stimulus payments	17,718 household who received rebates in 2001 or 2008	spending on nondurables and durables, plus subcategories	quarterly	reject	during six-months period in which payment was received, households spent 43 percent of payments on nondurable consumption in 2001, and 16 percent on total consumption in 2008	based on high income and high mortgage debt (wealthy hand-to-mouth consumers); evidence in favor
Parker (1999)	Caps on Social Security tax and changes in Social Security tax withholding; income increase and decrease	CEX interview survey (1980-1993)	133,820 observations on 57,051 households	spending on nondurables and durables, and different sub-categories	quarterly	reject	when a household's Social Security contributions fall so that income rises by \$1, nondurable consumption rises by 20 cents	based on age and liquid assets; weak evidence in favor
Parker, Souleles, Johnson, and McClelland (2013)	2008 Economic Stimulus Payments; income increase	CEX interview survey (2007-2008) with added questions on stimulus payments	17,478 household observations, of which 11,239 received economic stimulus payments	spending on nondurables and durables, plus subcategories	quarterly	reject	during three-months period in which payment was received, households increase their expenditures on nondurable goods by 12 to 39 percent of the payment, and on overall consumption by 50 to 90 percent	based on age, income, and liquid assets; some evidence based on income and age, not on liquid assets
Paxson (1993)	Seasonal income patterns in Thai agriculture; income increase and decrease	Thai Socio-economic Surveys (1975-1976, 1981, and 1986)	27,963 economically active (i.e., not retired) households that did not engage in forestry or fishing	spending on nondurables	monthly	cannot reject	spending patterns of farm and non-farm households are indistinguishable	no test

Study	Experiment	Data Source	Sample	Main Dependent Variable	Frequency of Data	PIH	Quantitative Reaction at Implementation	Liquidity Constraint
Scholnick (2013)	Last mortgage payment; disposable income increase	proprietary data set from a Canadian bank (2004-2006): credit card and mortgage accounts data	4,147 individuals who have payed off their mortgage or who have less than 1 year of mortgage payments left	credit card expenditures	monthly	reject	no quantitative interpretation reported; reaction of consumption to preannounced income increase is weaker the larger the income increase	based on households paying positive interest on credit card debt; also non liquidity constrained consumers react significantly
Shapiro (2005)	Food stamps; income increase and decrease	Continuing Survey of Food Intake by Individuals (1989–1991); Nationwide Food Consumption Survey (1987–1988): data on market value and nutritional characteristics of food eaten; survey conducted to document effects of Maryland's adoption of the Electronic Benefit Transfer system (1992-1993)	6,652 observations from surveyed individuals who receive food stamps	caloric intake	daily	reject	caloric intake declines by a statistically significant 0.40 percent per day after receipt of food stamps	no test
Shea (1995)	Unionized wage; income increase and decrease	PSID survey (1981-1986): data on food consumption matched with data on wage growth from union contracts	647 observations drawn from 285 households whose head is a union member and can be reasonably assigned to specific union	spending on food consumed at home and in restaurants, plus the bonus value of food stamps	annual	reject	a one percentage point increase in wage growth is associated with a 0.89 percentage point increase in food consumption	based on liquid assets and heterogeneous reactions to increases vs. decreases; mildly supportive of liquidity constraints in traditional measures; yet, reaction stronger to income decrease
Souleles (2002)	1981 Economic Recovery Tax Act (Reagan Tax Cuts); income increase	CEX interview survey (1982-1983)	2,399 household-quarter observations, head aged 24 to 64	spending on nondurables and total consumption	quarterly	reject	for each dollar increase in take-home pay, nondurable consumption rises by about two-thirds of a dollar	based on age, income, and liquid assets; no evidence in favor
Souleles (2000)	Paying for college; income decrease	CEX interview survey (1980-1993)	7,200 household observations with child aged 16–24, of which 1249 have positive college expenditure	spending on nondurables	quarterly	cannot reject	a one dollar decrease in income due to paying college tuition leads to an 8 cent increase (not decrease) in nondurable consumption	NA

Study	Experiment	Data Source	Sample	Main Dependent Variable	Frequency of Data	PIH	Quantitative Reaction at Implementation	Liquidity Constraint
Souleles (1999)	Income tax refunds; income increase	CEX interview survey (1980-1991)	4,121 observations on households receiving income tax refunds, head aged 24 to 64	spending on nondurables and total consumption	quarterly	reject	one dollar of refund receipt raises strictly nondurable consumption by 2.6 cents, and total consumption by 18 cents	based on liquid assets; some evidence in favor
Stephens (2008)	Final payment of a vehicle loan; disposable income increase	CEX interview survey (1984-2000)	4,583 observations on households who have an expiring vehicle loan	spending on nondurables, except for public transportation and gas and motor oil	quarterly	reject	a 10 percent increase in after-tax income increases non-durable consumption by 2.8 percent	based on age, liquid wealth, and maturity of expiring vehicle loan; evidence in favor based on age and liquid assets, but not based on maturity of prior loan
Stephens (2006)	Regular paycheck receipt; income increase and decrease	UK Family Expenditure Survey: two week diary of all expenditures (1986-1998)	12,827 households with a dependently employed monthly paid primary earner aged 25 to 59	total spending, spending on strict nondurables, food at home, and instant consumption goods	weekly	reject	instant consumption increases by 5 percent in week when households receive monthly paychecks	based on asset income; strong evidence in favor
Stephens (2003)	Receipt of Social Security check on "3rd of the Month"; income increase and decrease	CEX diary survey (1986-1996)	9,942 consumer units which contribute a total of 123,034 potential expenditure days	spending on instant consumption, food, and nondurables	daily	reject	households that receive at least 70 percent of income from Social Security increase instant consumption and food away from home by roughly 20 percent in week following arrival of Social Security check	no test
Stephens and Unayama (2011)	Change in frequency of Japanese public pension payments (February 1990); income increase and decrease	Japanese Family Income and Expenditure Survey (1986-1994): diary data on expenditures and income	2,503 retirees and employees before reform (pension paid once every 3 months), and 3,595 after reform (pension paid once every 2 months)	spending on nondurables and durables	monthly	reject	non-durable consumption increases by four percent in the month of check receipt, while strict non-durable and food consumption both increase by over two percent when checks are received	based on age, total net financial assets and demand deposits; no evidence in favor

Study	Experiment	Data Source	Sample	Main Dependent Variable	Frequency of Data	PIH	Quantitative Reaction at Implementation	Liquidity Constraint
Wilcox (1989)	changes in Social Security amounts; income increase	aggregate data on retail sales and personal consumption expenditures (1965-1985)		total retail sales, also divided in durable and non-durable good stores, and all commodities	monthly	reject	an increase in benefits by 10 percent increases total retail sales by 1.4 percent, with a 3 percent increase in durable goods sales (mostly cars)	no test
Liquidity constraint relaxation								
Abdallah and Lastrapes (2012)	1998 Texas constitutional amendment on home equity loans allows mortgages for non-housing expenditures	county (1992, 1997, 2002) and state (1992-2002) level data on retail sales; American Housing Survey (1994 and 2002); household-level data on second-lien equity homes	3006 counties / 45 states with complete data	retail sales (measured directly on county level, or as ratio of sales tax revenues to sales tax rates at state level)	annual (county level: 1992, 1997, 2002 only)	NA	real per capita spending of average Texas county increases by 2 to 4 percent due to relaxation of borrowing constraints	NA
Bertrand and Morse (2009)	2008 Economic Stimulus Payments; income increase	customers of a payday lending chain (March to September 2008)	881 active payday loan customers	payday loan take-up	weekly	NA	payday loan customers reduce borrowing on average by \$46 after receipt of the rebate check of on average \$600; frequency of borrowing also falls significantly	based on frequency of use of payday loans; no evidence in favor
Gross, Notowigdigdo, and Wang (2014)	2001 Federal Income Tax Rebates and 2008 Economic Stimulus Payments; income increase	data set compiled based on public access to Court Electronic Records system	all consumer bankruptcy filings in 81 out of 94 US courts (1998 to 2008)	Chapter 7 and Chapter 13 bankruptcy filings	weekly	NA	bankruptcies increase by 2% after 2001 rebates and 6% after 2008 rebates; also shift from Chapter 13 to Chapter 7 bankruptcies	based on income, share of subprime borrowers, and home ownership rate at ZIP code level; no evidence in favor
Leth-Petersen (2010)	1992 Danish credit market reform allows mortgages for non-housing expenditures	Danish public administrative registers: data on wealth, income, household composition, characteristics of dwelling (1987-1996)	63,613 households aged between 25 and 65 in 1991	imputed expenditures (income minus change in wealth)	annual	NA	credit constrained households with an equity to house value ratio of 0.5 or higher increase their expenditure by 1 to 3 percent in the years following change in law	based on liquid assets; some evidence in favor if equity to home value is high

Study	Experiment	Data Source	Sample	Main Dependent Variable	Frequency of Data	PIH	Quantitative Reaction at Implementation	Liquidity Constraint
Temporary price cuts								
Agarwal, Marwell, and McGranahan (2013)	Sales Tax Holidays	CEX diary survey (1997-2011); proprietary credit card transactions data from a large financial institution (2003, February 8 - October 20)	over 700,000 household-date observations from CEX diaries; over 10 million consumer-date observations from credit card data	spending on specific categories (esp. children's clothing), credit card transactions	daily	reject	sales tax holidays increase daily clothing spending by \$1.17 (i.e. 29% of daily household clothing spending); no significant reduction in spending before/after sales tax holiday on these categories, or during sales tax holidays on other categories	no test
Mian and Sufi (2012)	2009 Cars Allowance Rebate System ("Cash for Clunkers")	data on car purchases from R.L. Polk for US metropolitan or micropolitan statistical areas (2004-2010), augmented by data from census, Equifax Predictive Services, Federal Housing Finance Agency, BLS and IRS	957 metropolitan or micropolitan statistical areas	new car purchases	monthly	cannot reject	cities with lots of qualifying clunkers have significantly higher increase in car sales in months of program, but then reduce them, such that cumulative reponse over 12 months is zero	no test