The Pennsylvania State University

The Graduate School

College of the Liberal Arts

ESSAYS IN MACROECONOMIC THEORY AND POLITICAL ECONOMY

A Dissertation in

Economics

by

Sascha Claudius

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

August 2010

The dissertation of Sascha Claudius was reviewed and approved* by the following:

Edward J. Green
Professor of Economics
Dissertation Co-Advisor and Co-Chair of Committee

Herman J. Bierens
Professor of Economics
Dissertation Co-Advisor and Co-Chair of Committee

Alexander Monge-Naranjo
Assistant Professor of Economics

Burt Monroe
Associate Professor of Political Science

Vijay Krishna
Distinguished Professor of Economics
Director of Graduate Studies

*Signatures are on file in the Graduate School.

Abstract

This thesis consists of three chapters. The first chapter investigates how the functional operator used in Kocherlakota (1996) is linked to the game underlying the model. Using the machinery developed by Abreu, Pearce, and Stacchetti (1990) (APS) I solve for the set of equilibrium value profiles directly. Consider a function f whose hypograph is self-generating and apply the APS set operator B to the hypograph. The main result states that the functional operator used in Kocherlakota (1996) maps the function f to the upper boundary of the resulting set.

The second chapter tackles various issues in the empirical literature investigating the claim that economic development leads to political development, i.e., higher income causes democracy. All democracy data are ordinal, i.e., ordered and discrete. In contrast to Barro (1999) and Acemoglu, Johnson, Robinson, and Yared (2008), who treat the democracy data as continuous, I develop and estimate a dynamic model where the ordinal democracy variable has the Markov property. As pointed out by Robinson (2006), most work in the literature fails to take into account that democracy might have an effect on income. Consequently, I test whether income has an effect on democracy and whether democracy has an effect on income. The effects between income and democracy are dynamic as they unfold over time. This leads me to employ the standard econometric causality concept in time series analysis, Granger causality. Mosconi and Seri (2006) have extended the concept of Granger causality to binary processes. Using a similar Markov approach allows me to apply their definition of Granger causality in my econometric model. I also show that GDP per capita is a unit root process for almost all countries and that failing to take this into account leads to model misspecification and misleading results. My empirical analysis suggests the following. First, higher income Granger causes democracy for full democracies, which means that a high income helps to stabilize democracies. Second, higher income Granger causes democracy for medium democracies implying that income can have a positive effect on regime type where there is already a somewhat favorable environment. Third, I

find that Granger causality runs from democracy to income meaning that income and democracy form a bidirectional feedback system.

The third chapter investigates empirically the effect of extending the franchise to women. The underlying question is why women had to fight for their right to vote. The hypothesis is that richer and more educated women had a higher turnout at the polls than poorer and less educated women. Using data on demographic composition, election results and voter registration for Pittsburgh I estimate female voter turnout and female voter participation across different social groups and the effect of women suffrage on election results. I find that female voter turnout and female voter participation was higher in the richer and more educated parts of town. However, I cannot find an effect of the enactment women suffrage on the election results.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am most grateful and indebted to my thesis advisors, Herman J. Bierens and Edward J. Green, for their advice, guidance and support. I would like to thank my wife for her support and encouragement.

Dedication

For my dear parents, siblings, and my beloved wife who filled my life with happiness.

Chapter 1

A Note on the Game-Theoretic Foundations of Infinite Horizon Environments with Limited Commitment

1.1 Introduction

Limited commitment as a restriction to individual behavior has attracted strong interest in the macroeconomic theory literature. According to Feller (1950) a field is successful if it can provide intuition, applications and results. First, it is clear that contracts among governments cannot be enforced. Also, at least in Western democracies courts are unwilling to enforce contracts that regulate an individual's behavior over 40 years. Second, the implications of the individual and institutional inability to commit to long term contracts have been explored for asset pricing, inequality, and international lending.[1] Third, Thomas and Worrall (1988) and Kocherlakota (1996) have developed a procedure that enables researchers to produce sharp results at a very small technical expense. The idea is as follows.

There are at least two parties engaged in a long term relationship. A stable contract needs to account for the possibility that each party can suspend the arrangement at any point in time. This means that there is complex interaction over time that can be formalized using game theory. Thomas and Worrall (1988) and Kocherlakota (1996) choose to avoid setting up the game explicitly and use instead lifetime values and outcomes yielding a massive reduction in complexity. In particular, Kocherlakota's approach works as follows.

First, use a function to describe the agent's lifetime values associated with the efficient equilibrium outcomes. Second, show that this function satisfies a particular functional equation. Third, characterize the solutions to this functional equation. This functional equation can be viewed as a functional operator. In what follows I call this particular functional equation the functional operator T. In this sense, the second step shows that the function in the first step is a fixed point of the operator T. The only justification for the use of operator T is that it works. The relationship between this operator and the economic environment is typically established via intuition. This note investigates how the functional operator T is related to the explicit game underlying Kocherlakota's model. The objective is to link this operator to the equilibrium conditions given by the APS framework.

I start by setting up Kocherlakota' model explicitly as a repeated game. Using the machinery developed by APS I solve for the equilibrium values directly. In particular, I define admissibility and prove the factorization and the self-generation

[1] See Alvarez and Jermann (2001), Krueger and Perri (2006) and Thomas and Worrall (1994).

result for this environment. This is necessary even though the environment is a special case of the environment originally considered in Abreu, Pearce, and Stacchetti (1990). The reason is that the equilibrium concept *sequential equilibrium* used by APS is much more restrictive than the equilibrium concept *subgame perfectness* used by Kocherlakota.[2] These intermediary results enable me to derive the main result. Consider a function f whose hypograph is self-generating and apply the APS set operator B to this hypograph. Call the upper boundary of the resulting set g. The main result states that Kocherlakota's operator T maps the function f to the function g.

The outline of the paper is as follows: First, I setup the environment and define the equilibrium concept. Then, the APS framework is established. The main result is derived in the final section.

1.2 Environment

Time is discrete and indexed by $t = 1, 2, ...$ There are two infinitely lived agents A and B and N states of the world, $\Omega = \{{}^1\omega, ..., {}^N\omega\}$. The distribution of states is i.i.d. and the probability of state ${}^k\omega$ occurring is given by $\pi({}^k\omega) = \pi_k$ with $\sum_{k=1}^N \pi_k = 1$.[3] A state in period t is denoted by ω_t and a history of states in t is given by $\omega^t = (\omega_1, ..., \omega_t)$. The aggregate endowment in period t with date$-t$ consumption goods is 1. Denote by $y^i(\omega)$ i's endowment if the current state is ω and assume $y^A(\omega) + y^B(\omega) = 1$ with $y^i(\omega) \in [0, 1]$. Period utility $u : [0, 1] \rightarrow R_+$ is weakly increasing. At the beginning of period t each agent i learns the current state ω_t and subsequently transfers tr_t^i to agent $j \neq i$. Denote by $tr^{it} =$

[2] In particular, formulating Kocherlakota's environment in the language of Abreu, Pearce, and Stacchetti (1990) leads to a failure of their assumption (A3).

[3] The distribution of the shocks is not relevant for the results.

$\left(tr_1^i, ..., tr_t^i\right)$ i's past transfers to j. Define for all $t \geq 1$ a history of past events $h^t \equiv (\omega^t, tr^{At}, tr^{Bt}) \in H^t \equiv \Omega^t \times [0,1]^t \times [0,1]^t$ and set $H^0 = \{\emptyset\}$. The set of all histories is given by $\mathcal{H} = \bigcup_{t=0}^{\infty} H^t$. A strategy for player i is defined by $\tau_i : \Omega \times \mathcal{H} \rightarrow [0,1]$. Define the (non-commutative) operator $*$ by

$$
\begin{aligned}
\tilde{h}^t * \hat{h}^d &= \left((\tilde{\omega}_1, ..., \tilde{\omega}_t), (\tilde{tr}_1^A, ..., \tilde{tr}_t^A), (\tilde{tr}_1^B, ..., \tilde{tr}_t^B)\right) \\
&* \left((\hat{\omega}_1, ..., \hat{\omega}_d), (\hat{tr}_1^A, ..., \hat{tr}_d^A), (\hat{tr}_1^B, ..., \hat{tr}_d^B)\right) \\
&= \left((\tilde{\omega}_1, ..., \tilde{\omega}_t, \hat{\omega}_1, ..., \hat{\omega}_d), (\tilde{tr}_1^A, ..., \tilde{tr}_t^A, \hat{tr}_1^A, ..., \hat{tr}_d^A),\right. \\
&\quad \left.(\tilde{tr}_1^B, ..., \tilde{tr}_t^B, \hat{tr}_1^B, ..., \hat{tr}_d^B)\right) \\
&= \left((\omega_1, ..., \omega_t, \omega_{t+1}, ..., \omega_{t+d}), (tr_1^A, ..., tr_t^A, tr_{t+1}^A, ..., tr_{t+d}^A),\right. \\
&\quad \left.(tr_1^B, ..., tr_t^B, tr_{t+1}^B, ..., tr_{t+d}^B)\right).
\end{aligned}
$$

A continuation strategy $\tau_i|_{h^t}$ after history h^t is defined as $\tau_i|_{h^t}(\omega, h^d) = \tau_i(\omega, h^t * h^d)$ for all $h^d \in \mathcal{H}$. A path of states $\omega^t = (\omega_1, ..., \omega_t)$ and a strategy profile $\tau = (\tau_A, \tau_B)$ induce individual transfers recursively as follows:

$$
\begin{aligned}
tr_1^i(\omega^1, \tau) &= \tau_i(\omega_1, h^0) \qquad & h^1 &= \left(\omega_1, tr_1^A(\omega^1, \tau), tr_1^B(\omega^1, \tau)\right) \\
tr_s^i(\omega^s, \tau) &= \tau_i(\omega_s, h^{s-1}) \qquad & h^s &= h^{s-1} * \left(\omega_s, tr_s^A(\omega^s, \tau), tr_s^B(\omega^s, \tau)\right)
\end{aligned}
$$

for all $1 < s \leq t$ with ω^s being the projection of ω^t into its first s dimensions. E_t denotes the expectation over future states starting with the unknown states in period t. Agent i maximizes average discounted utility. Agent i's sequential lifetime value in period $t+1$ after learning the current state ω_{t+1} given history

$h^t \in H^t$ and strategy profile $\tau|_{h^t}$ is given by,

$$V^i(\tau|_{h^t})(\omega_{t+1}) \quad = \quad \frac{1-\beta}{\beta} E_{t+2} \sum_{s=t+1}^{\infty} \beta^{s-t} u\big(y^i(\omega_s) - tr_s^i(\omega^s, \tau|_{h^t}) + tr_s^j(\omega^s, \tau|_{h^t})\big).$$

Agent i's recursive lifetime utility in period $t+1$ given current state ω_{t+1} and strategy profile $\tau|_{h^t}$ is given by

$$\begin{aligned} V^i(\tau|_{h^t})(\omega_{t+1}) \quad &= \quad (1-\beta)u\big(y^i(\omega_{t+1}) - tr_{t+1}^i(\omega^{t+1}, \tau|_{h^t}) + tr_{t+1}^j(\omega^{t+1}, \tau|_{h^t})\big) \\ &+ \quad \beta \sum_{\omega_{t+2} \in \Omega} V^i(\tau|_{h^t * h^1})(\omega_{t+2}) \pi(\omega_{t+2}). \end{aligned}$$

Agent i's lifetime value prior to learning the current state is given by $V^i(\tau|_{h^t}) = \sum_{\omega_{t+1} \in \Omega} V^i(\tau|_{h^t})(\omega_{t+1}) \pi(\omega_{t+1})$.

1.3 Equilibrium

So far, an agent is allowed to make infeasible transfers. Agent i's strategy τ_i is said to be *feasible* if for all $(\omega, h^t) \in \Omega \times \mathcal{H}$, $0 \leq \tau_i(\omega, h^t) \leq y^i(\omega)$. The natural equilibrium concept for an environment with two-sided limited commitment is subgame perfectness. It is most convenient to use Nash equilibrium to define subgame perfectness.

Definition 1 *The strategy profile τ constitutes a* Nash equilibrium[4] *if for all agents i and for all feasible strategies $\hat{\tau}_i$*

$$V^i(\hat{\tau}_i, \tau_j) \leq V^i(\tau).$$

[4] Strictly speaking it is a Bayesian Nash equilibrium.

This definition applies to continuation strategies as well. The concept of Nash equilibrium leads to the concept of subgame perfectness.[5]

Definition 2 *The strategy profile* τ *constitutes a* subgame-perfect equilibrium (SPE) *if for all* $(\omega_{t+1}, h^t) \in \Omega \times \mathcal{H}$, $\tau|_{h^t}$ *is a Nash equilibrium.*

As before, this definition applies to continuation strategies as well. Note the following. First, the one-deviation property holds for SPE strategy profiles.[6] Second, autarchy refers to the strategy profile specifying $\tau_i(\omega, h^t) = 0$ for all i, ω, h^t. Let $V^i_{aut}(\omega)$ denote the associate value for agent i in state ω. Since zero transfers is the unique Nash equilibrium of the stage game, autarchy is an SPE in the repeated game. Furthermore, since autarchy delivers the minmax payoff, autarchy is the worst SPE.[7]

The main result of this paper is shown using the machinery developed in Abreu, Pearce, and Stacchetti (1990) (APS). One of their fundamental ideas is to work with lifetime values instead of strategies. This approach makes extensive use of the set of ex-ante SPE lifetime utility profiles,

$$\begin{aligned} W \;=\; & \big\{(v^A, v^B) \in R^2_+ \mid \text{ there exists an SPE strategy profile } \tau|_{h^t} \text{ such} \\ & \text{that } v^A = V^A(\tau|_{h^t}) \text{ and } v^B = V^B(\tau|_{h^t})\big\}. \end{aligned}$$

A consequence of APS working with lifetime values instead of strategies is that their game becomes static. The reduction in complexity is achieved by working

[5] The set of Nash equilibrium (NE) allocations is identical to the set of SPE allocations. Every SPE allocation is supported by the credible threat to play the worst SPE in case of a deviation. The worst SPE is autarky. Every NE allocation is supported by the threat to minmax the deviating player forever. Autarky yields the minmax payoff profile. Van Damme (1987) proves that if the worst NE of the stage game yields the minmax payoff every NE allocation is also SPE.

[6] The proof for the deterministic case can be found in Fudenberg and Tirole (1991), p.110. The present setup is a stochastic version of what Fudenberg and Tirole (1991) call an infinite horizon multi-stage game with observed actions. See also Mailath and Samuelson (2006), p.25.

[7] This remark is made by Mailath and Samuelson (2006), p.211.

with current actions and continuation values. This raises the question how an equilibrium of the static game relates to an equilibrium of the dynamic game. The next lemma helps to answer this question. The proof is in appendix A.

Lemma 1 (Decomposition) *A feasible strategy profile $\tau|_{h^t}$ constitutes an SPE if and only if*

1. *the feasible continuation strategy profile $\tau|_{h^t * h^1}$ constitutes an SPE strategy profile for all $h^1 \in H^1$*

2. *for all agents i, all $\omega_{t+1} \in \Omega$ and all feasible $\hat{\tau}_i|_{h^t}$*

$$\begin{aligned} & (1-\beta)u\big(y^i(\omega_{t+1}) - tr^i_{t+1}(\omega^{t+1}, \tau|_{h^t}) + tr^j_{t+1}(\omega^{t+1}, \tau|_{h^t})\big) \\ + \quad & \beta \sum_{\omega_{t+2}\in\Omega} V^i(\tau|_{h^{t+1}})(\omega_{t+2})\pi(\omega_{t+2}) \\ \geq \quad & (1-\beta)u\big(y^i(\omega_{t+1}) - tr^i_{t+1}(\omega^{t+1}, (\hat{\tau}_i|_{h^t}, \tau_j|_{h^t})) + tr^j_{t+1}(\omega^{t+1}, (\hat{\tau}_i|_{h^t}, \tau_j|_{h^t}))\big) \\ + \quad & \beta \sum_{\omega_{t+2}\in\Omega} V^i(\hat{\tau}_i|_{\hat{h}^{t+1}}, \tau_j|_{\hat{h}^{t+1}})(\omega_{t+2})\pi(\omega_{t+2}) \end{aligned}$$

with

$$\begin{aligned} h^{t+1} &= h^t * \big(\omega_{t+1}, tr^i_{t+1}(\omega^{t+1}, \tau|_{h^t}), tr^j_{t+1}(\omega^{t+1}, \tau|_{h^t})\big) \\ \hat{h}^{t+1} &= h^t * \big(\omega_{t+1}, tr^i_{t+1}(\omega^{t+1}, (\hat{\tau}_i|_{h^t}, \tau_j|_{h^t})), tr^j_{t+1}(\omega^{t+1}, (\hat{\tau}_i|_{h^t}, \tau_j|_{h^t}))\big). \end{aligned}$$

Lemma 1 states that a strategy profile is subgame perfect if and only if it can be decomposed into an optimal current action and subgame perfect continuation strategy profiles. Next, I develop the APS framework.

1.4 APS

The objective of this section is to show that the set of value profiles supported by subgame perfect strategy profiles is a fixed point of the set operator B.

First, I define admissibility. A reward function has the functional form $p_\omega : [0,1]^2 \to R^2$ with $p_\omega(tr_\omega^A, tr_\omega^B) = [p_\omega^A(tr_\omega^A, tr_\omega^B), p_\omega^B(tr_\omega^A, tr_\omega^B)]$. If the current period's state is ω $p_\omega^i(tr_\omega^A, tr_\omega^B)$ is agent i's expected lifetime value tomorrow conditional on agent j transferring tr_ω^j to agent i today. The set of possible lifetime value profiles is given by $D = [u(0), u(1)]^2$.

Definition 3 *Let $L \subseteq D$. An object $T = \{(tr_\omega^A, tr_\omega^B, p_\omega)\}_{\omega\in\Omega}$ is said to be* admissible *with respect to L if for all $\omega \in \Omega$ and all i*

1. $tr_\omega^i \in [0, y^i(\omega)]$
2. $p_\omega : [0,1]^2 \to L$
3. *for all* $\hat{tr}^i \in [0, y^i(\omega)]$

$$\begin{aligned} & (1-\beta)u(y^i(\omega) - tr_\omega^i + tr_\omega^j) + \beta p_\omega^i(tr_\omega^i, tr_\omega^j) \\ \geq\ & (1-\beta)u(y^i(\omega) - \hat{tr}^i + tr_\omega^j) + \beta p_\omega^i(\hat{tr}^i, tr_\omega^j). \end{aligned}$$

The first condition requires that the transfers must be feasible while the second condition places restrictions on the continuation values. The last condition requires incentive compatibility. Henceforth, we say T *supports* (w^A, w^B) if

$$w^i = \sum_{\omega\in\Omega} \pi(\omega)\big[(1-\beta)u(y^i(\omega) - tr_\omega^i + tr_\omega^j) + \beta p_\omega^i(tr_\omega^i, tr_\omega^j)\big].$$

This leads to the definition of the set operator B.

Definition 4 *Define an operator* $B : 2^D \to 2^D$ *by*

$$\begin{aligned} B(L) &= \{(w^A, w^B) \mid \textit{there exists an admissible } T \textit{ with respect to } L \\ &\quad \textit{supporting } (w^A, w^B)\}. \end{aligned}$$

Loosely speaking, the B operator collects all lifetime profiles that are feasible, incentive compatible and have continuation value profiles in L. The next three lemmas establish that that the sets of subgame perfect continuation values W is the largest fixed point of the set operator B. The first lemma is technical. All proofs are in appendix A.

Lemma 2 (Monotonicity of B). *If* $L \subseteq \tilde{L}$, *then* $B(L) \subseteq B(\tilde{L})$.

The next definition introduces the key concept. A set is self-generating if every value profile can be supported by a current action and a continuation value profile such that the continuation value profile is an element of the very same set.

Definition 5 L *is said to be* self-generating, *if* $L \subseteq B(L)$.

The below lemma confirms a simple intuition − subgame perfect value profiles can be supported with subgame perfect continuation value profiles.

Lemma 3 (Factorization) W *is self-generating.*

The next lemma reverses the question: Is every element of a self-generating set a subgame perfect value profile?

Lemma 4 (Self-generation) *If* L *is self-generating, then* $L \subseteq W$.

To summarize, we have shown the following. First, by Lemma 3 $W \subseteq B(W)$. This implies by Lemma 2 $B(W) \subseteq B(B(W))$. Hence, $B(W)$ is self-generating. This in turn implies by Lemma 4 $B(W) \subseteq W$ yielding $W = B(W)$. Since any fixed point FP must be self-generating, by Lemma 4 $FP \subseteq W$, so W is the largest fixed point of $B(\cdot)$. We conclude the section by showing that the operator B preserves compactness and W is compact.

Lemma 5 (Compactness) *If L is compact, then $B(L)$ is compact.*

Note that $W \subseteq D = [u(0), u(1)]^2$. By monotonicity, $B(W) \subseteq B(D)$ and hence $W \subseteq B(D)$. Similarly one shows that $W \subseteq B^n(D)$. Since $B(D) \subseteq D$, $B^{n+1}(D) \subseteq B^n(D)$. Finally because $W = B(W)$, $\lim_{n\to\infty} B^n(D) = W$ and W is compact.

1.5 The Relationship between APS and Kocherlakota's operator T

This section provides the link between the set of equilibrium value profiles W given by the APS set operator B and Kocherlakota's operator T. The idea is as follows. Let V_{max} be the highest equilibrium value any player can obtain. Since W is compact this value exists. Consider figure 1.1 depicting the equilibrium value set W on page 11.

The area below the function f is the hypograph of f. Assume this hypograph is self-generating and apply the set operator B to this hypograph. Call the upper boundary of the resulting set g. The main result says that Kocherlakota's operator T maps the function f to the function g. Below I make this argument formally.

First, define the function space $C = \{f : [V_{aut}, V_{max}] \to [V_{aut}, V_{max}]\}$. For $f \in C$ define its hypograph as $hyp(f) = \{(x, y) | x \in [V_{aut}, V_{max}], V_{aut} \le y \le f(x)\}$.

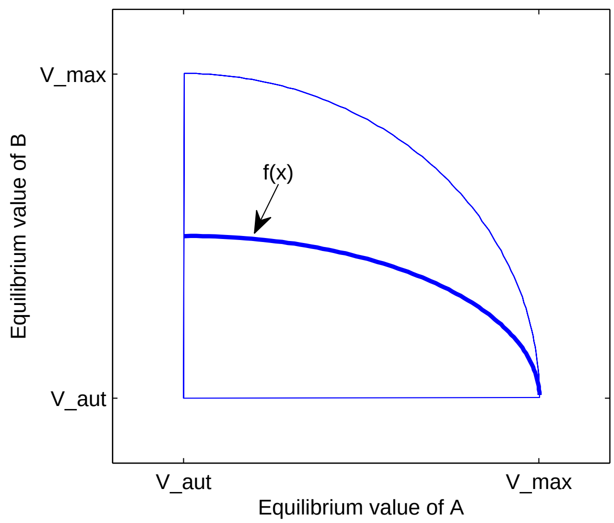


Figure 1.1. The Equilibrium Value Set W.

Applying the set operator B to the hypograph of f yields

$$B\big(hyp(f)\big) = \{(w^A, w^B)| \text{ there exists an admissible } \{(tr_\omega^A, tr_\omega^B, p_\omega)\}_{\omega\in\Omega} \text{ with respect to } hyp(f) \text{ supporting } (w^A, w^B)\}.$$

Next, define

$$\widetilde{C} = \{f : [V_{aut}, V_{max}] \rightarrow [V_{aut}, V_{max}] \mid hyp(f) \text{ is self-generating}\}.$$

Note the following. First, the function describing the efficient frontier of W is an element of $\widetilde{C}$. Hence, $\widetilde{C}$ is non-empty. Second, the fact that $hyp(f)$ is self-generating in conjunction with the monotonicity of B yields $hyp(f) \subseteq B\big(hyp(f)\big) \subseteq W$. Third, since $hyp(f) \subseteq B\big(hyp(f)\big)$, we have $[V_{aut}, V_{max}] \times \{V_{aut}\} \subseteq B\big(hyp(f)\big)$. In other words, for every $w^A \in [V_{aut}, V_{max}]$ there exists an $w^B \in [V_{aut}, V_{max}]$ such that $(w^A, w^B) \in B\big(hyp(f)\big)$.

The objective is to link the upper boundary of $B(hyp(f))$ to Kocherlakota's operator T. The next lemma shows that every element of this upper boundary can be supported by some vector $\{(tr^A_\omega, tr^B_\omega, w^A_\omega, w^B_\omega)\}_{\omega\in\Omega}$ subject to feasibility, incentive compatibility and the continuation value profiles being an element of $hyp(f)$. The major simplification is that instead of considering reward functions it is sufficient to consider a vector of continuation value profiles in case of compliance by the agents. The reason is that every $(w^A, w^B) \in B\big(hyp(f)\big)$ is an element of W and in order to support elements of W it is sufficient to punish deviations with the worst SPE continuation value, namely autarchy. The proof is in appendix A.

Lemma 6 *Let $f \in \widetilde{C}$. For $(w^A, w^B) \in B\big(hyp(f)\big)$ it is sufficient find a vector $\{(tr^A_\omega, tr^B_\omega, w^A_\omega, w^B_\omega)\}_{\omega\in\Omega}$ supporting (w^A, w^B) such that for all $\omega \in \Omega$ and all i*

1. $tr^i_\omega \in [0, y^i(\omega)]$
2. $(w^A_\omega, w^B_\omega) \in hyp(f)$
3. *for all* $\hat{tr}^i \in [0, y^i(\omega)]$

$$(1-\beta)u(y^i(\omega) - tr^i_\omega + tr^j_\omega) + \beta w^i_\omega \geq (1-\beta)u(y^i(\omega) - \hat{tr}^i + tr^j_\omega) + \beta V_{aut}.$$

Now, I provide the main argument. Let $f \in \widetilde{C}$ and denote the upper boundary of $B\big(hyp(f)\big)$ by $UB_{B(hyp(f))}$. $UB_{B(hyp(f))}$ is given by

$$UB_{B(hyp(f))}(w^A) = \max \hat{w}^B \text{ s.t. } (w^A, \hat{w}^B) \in B\big(hyp(f)\big).$$

Next, I argue that $UB_{B(hyp(f))} \in \widetilde{C}$. First, since $[V_{aut}, V_{max}] \times \{V_{aut}\} \subseteq B\big(hyp(f)\big) \subseteq W$, $[V_{aut}, V_{max}]$ can serve as the domain of $UB_{B(hyp(f))}$. Second, since $B\big(hyp(f)\big) \subseteq$

W the image of $UB_{B(hyp(f))}$ is a subset of $[V_{aut}, V_{max}]$. Third, since $hyp(f)$ is compact and B preserves compactness by Lemma 5, the maximum is well-defined and unique. Together, these three points imply that $UB_{B(hyp(f))} : [V_{aut}, V_{max}] \to [V_{aut}, V_{max}]$ is well-defined.

Fourth, $hyp\big(UB_{B(hyp(f))}\big) = B\big(hyp(f)\big)$. Since $hyp(f)$ is self-generating, $hyp(f) \subseteq B\big(hyp(f)\big)$. By monotonicity, $B\big(hyp(f)\big) \subseteq B\big(B\big(hyp(f)\big)\big)$. This implies $B\big(hyp(f)\big)$ is self-generating, implying $hyp\big(UB_{B(hyp(f))}\big)$ is self-generating. Hence, $UB_{B(hyp(f))} \in \widetilde{C}$. This establishes that that the following operator is well-defined: $\widetilde{T} : \widetilde{C} \to \widetilde{C}$ with

$$\widetilde{T}(f)(w^A) = \max\ \hat{w}^B \text{ s.t. } (w^A, \hat{w}^B) \in B\big(hyp(f)\big).$$

Using Lemma 6 this can be rewritten as

$$\begin{aligned}
\widetilde{T}(f)(w^A) \quad = \quad & \max_{(tr^A_\omega, tr^B_\omega, w^A_\omega, w^B_\omega)_\omega} \sum_{\omega\in\Omega} \pi(\omega)\big[(1-\beta)u(y^B(\omega) - tr^B_\omega + tr^A_\omega) + \beta w^B_\omega\big] \\
\text{s.t.} \quad \bullet \quad & \sum_{\omega\in\Omega} \pi(\omega)\big[(1-\beta)u(y^A(\omega) - tr^A_\omega + tr^B_\omega) + \beta w^A_\omega\big] = w^A \\
\bullet \quad & 0 \le tr^i_\omega \le y^i(\omega) \\
\bullet \quad & \quad (1-\beta)u(y^A(\omega) - tr^A_\omega + tr^B_\omega) + \beta w^A_\omega \\
& \ge (1-\beta)u(y^A(\omega) - tr^A + tr^B_\omega) + \beta V_{aut} \\
\bullet \quad & \quad (1-\beta)u(y^B(\omega) - tr^B_\omega + tr^A_\omega) + \beta w^B_\omega \\
& \ge (1-\beta)u(y^B(\omega) - tr^B + tr^A_\omega) + \beta V_{aut} \\
\bullet \quad & (w^A_\omega, w^B_\omega) \in hyp(f).
\end{aligned}$$

Now, suppose $w^B_\omega < f(w^A_\omega)$ for some $\omega \in \Omega$. Then $\widetilde{T}(f)(w^A)$ cannot be a maximum, so write

$$\begin{aligned}
\widetilde{T}(f)(w^A) &= \max_{(tr^A_\omega, tr^B_\omega, w^A_\omega)_\omega} \sum_{\omega\in\Omega} \pi(\omega)\big[(1-\beta)u(y^B(\omega) - tr^B_\omega + tr^A_\omega) + \beta f(w^A_\omega)\big] \\
\text{s.t.} \quad & \bullet \ \sum_{\omega\in\Omega} \pi(\omega)\big[(1-\beta)u(y^A(\omega) - tr^A_\omega + tr^B_\omega) + \beta w^A_\omega\big] = w^A \\
& \bullet \ 0 \le tr^i_\omega \le y^i(\omega) \\
& \bullet \quad (1-\beta)u(y^A(\omega) - tr^A_\omega + tr^B_\omega) + \beta w^A_\omega \\
& \ge (1-\beta)u(y^A(\omega) - tr^A + tr^B_\omega) + \beta V_{aut} \\
& \bullet \quad (1-\beta)u(y^B(\omega) - tr^B_\omega + tr^A_\omega) + \beta f(w^A_\omega) \\
& \ge (1-\beta)u(y^B(\omega) - tr^B + tr^A_\omega) + \beta V_{aut} \\
& \bullet \ (w^A_\omega, f(w^A_\omega)) \in hyp(f).
\end{aligned}$$

Note that the last condition is equivalent to the condition $w^A_\omega \in [V_{aut}, V_{max}]$. If $(w^A_\omega, f(w^A_\omega)) \in hyp(f)$ then since $f \in \widetilde{C}$, $w^A_\omega \in [V_{aut}, V_{max}]$. Conversely, if $w^A_\omega \in [V_{aut}, V_{max}]$ then $(w^A_\omega, f(w^A_\omega)) \in hyp(f)$ by definition of $hyp(f)$. So write,

$$\begin{aligned}
\widetilde{T}(f)(w^A) &= \max_{(tr^A_\omega, tr^B_\omega, w^A_\omega)_\omega} \sum_{\omega\in\Omega} \pi(\omega)\big[(1-\beta)u(y^B(\omega) - tr^B_\omega + tr^A_\omega) + \beta f(w^A_\omega)\big] \\
\text{s.t.} \quad & \bullet \ \sum_{\omega\in\Omega} \pi(\omega)\big[(1-\beta)u(y^A(\omega) - tr^A_\omega + tr^B_\omega) + \beta w^A_\omega\big] = w^A \\
& \bullet \ 0 \le tr^i_\omega \le y^i(\omega) \\
& \bullet \quad (1-\beta)u(y^A(\omega) - tr^A_\omega + tr^B_\omega) + \beta w^A_\omega \\
& \ge (1-\beta)u(y^A(\omega) - tr^A + tr^B_\omega) + \beta V_{aut} \\
& \bullet \quad (1-\beta)u(y^B(\omega) - tr^B_\omega + tr^A_\omega) + \beta f(w^A_\omega) \\
& \ge (1-\beta)u(y^B(\omega) - tr^B + tr^A_\omega) + \beta V_{aut} \\
& \bullet \ w^A_\omega \in [V_{aut}, V_{max}].
\end{aligned}$$

This is essentially Kocherlakota's operator T. Redefine $v(\cdot) = (1-\beta)u(\cdot)$, let agent B (A) become agent 2 (1), define $c_s^i = y^i(\omega) - tr_\omega^i + tr_\omega^j$ for $\omega = s$, $x_s = w_\omega^A$ for $\omega = s$, $\chi = w^A$ and assume all consumption values are feasible. This yields Kocherlakota's operator T

$$
\begin{aligned}
T(f)(\chi) \quad &= \quad \max_{(c_s^1, c_s^2, x_s)_{s=1}^S} \sum_{s=1}^{S} \pi_s [v(c_s^2) + \beta f(x_s)] \\
\text{s.t.} \quad & \quad \sum_{s=1}^{S} \pi_s [v(c_s^1) + \beta x_s] = \chi \\
& \quad v(c_s^1) + \beta x_s \geq v(y_s^1) + \beta V_{aut} \\
& \quad v(c_s^2) + \beta f(x_s) \geq v(y_s^2) + \beta V_{aut} \\
& \quad x_s \in [V_{aut}, V_{max}].
\end{aligned}
$$

This proves the main result.

Theorem 1 *Kocherlakota's operator T has the following property:*

$$T : \widetilde{C} \rightarrow \widetilde{C} \text{ with } T(f) = UB_{B(hyp(f))}.$$

1.6 Appendix A

Proof of Lemma 1 The proof follows from two observations. Condition 2 states that $\tau|_{h^t}$ is a Nash equilibrium for all (ω_{t+1}, h^0). Condition 1 states that for all (ω_{t+s}, h^{s-1}) with $s \geq 2$, $\tau|_{h^t * h^{s-1}}$ is a Nash equilibrium. Hence, $\tau|_{h^t}$ is a Nash equilibrium for all $(\omega_{t+d}, h^d) \in \Omega \times \mathcal{H}$. □

Proof of Lemma 2 Let $(w^A, w^B) \in B(L)$. Then there exists an admissible $T = \{(tr^A_\omega, tr^B_\omega, p_\omega)\}_{\omega \in \Omega}$ with respect to L supporting (w^A, w^B). The idea is to show that T is also admissible with respect to $\tilde{L}$. Since $L \subseteq \tilde{L}$ and $p_\omega(\cdot,\cdot) \in L$, $p_\omega(\cdot,\cdot) \in \tilde{L}$. This takes care of the second condition. If the third condition holds for L it also holds for $\tilde{L}$. Hence, T is admissible with respect to $\tilde{L}$ and $(w^A, w^B) \in B(\tilde{L})$. This yields $B(L) \subseteq B(\tilde{L})$. □

Proof of Lemma 3 (By construction). Let $(w^A, w^B) \in W$. Then there exists a subgame perfect strategy profile $\tau|_{h^t}$ such that $(w^A, w^B) = (V^A(\tau|_{h^t}), V^A(\tau|_{h^t}))$. Hence,

$$\begin{aligned} w^i &= V^i(\tau|_{h^t}) = \sum_{\omega_{t+1} \in \Omega} \Big[(1-\beta) u\big(y^i(\omega_{t+1}) - tr^i_{t+1}(\omega^{t+1}, \tau|_{h^t}) \\ &+ tr^j_{t+1}(\omega^{t+1}, \tau|_{h^t})\big) + \beta V^i(\tau|_{h^t * h^1})\Big] \pi(\omega_{t+1}) \end{aligned}$$

with $h^1 = \big(\omega_{t+1}, tr^A_{t+1}(\omega^{t+1}, \tau|_{h^t}), tr^B_{t+1}(\omega^{t+1}, \tau|_{h^t})\big)$. The idea is to construct a T which is admissible with respect to W and supports (w^A, w^B). To this end, define first for all $\omega \in \Omega$ $tr^i_\omega = tr^i_{t+1}(\omega^{t+1}, \tau|_{h^t})$ with $\omega = \omega_{t+1}$. Second, define for all $\omega \in \Omega$ a function $p_\omega : [0,1]^2 \to W$ with $p_\omega(tr^A_\omega, tr^B_\omega) = \big(V^A(\tau|_{h^t * h^1}), V^B(\tau|_{h^t * h^1})\big)$. By Lemma 1, $\tau|_{h^t * h^1}$ constitutes a subgame perfect strategy profile implying $p_\omega(tr^A_\omega, tr^B_\omega) \in W$. Next, for all $\widetilde{tr}^i \neq tr^i_\omega \in [0, y^i(\omega)]$ there exists a strategy $\widetilde{\tau}_i|_{h^t}$ specifying $\widetilde{\tau}_i|_{h^t}(\omega, h^0) = \widetilde{tr}^i$ and $\widetilde{\tau}_i|_{h^t * h^d} = \tau_i|_{h^t * h^d}$ for all $d > 0$ and $h^d \in \mathcal{H}$. Define

$p_\omega(\widetilde{tr}^i, tr^j_\omega) = \big(V^i(\widetilde{\tau}_i|_{h^t * \widetilde{h}}, \tau_j|_{h^t * \widetilde{h}}),$ $V^j(\widetilde{\tau}_i|_{h^t * \widetilde{h}}, \tau_j|_{h^t * \widetilde{h}})\big)$ with $\widetilde{h} = \big(\omega, \widetilde{tr}^i, tr^j_\omega\big)$. Since $(\widetilde{\tau}_i|_{h^t * \widetilde{h}}, \tau_j|_{h^t * \widetilde{h}}) = \tau|_{h^t * \widetilde{h}}$ and $\tau|_{h^t * \widetilde{h}}$ constitutes a subgame perfect strategy profile by Lemma 1, $p_\omega(\widetilde{tr}^i, tr^j_\omega) \in W$. Assign an arbitrary element of W to all other elements in the support of $p_\omega(\cdot,\cdot)$ so that $p_\omega(\cdot,\cdot)$ is well-defined. It remains to show that the constructed $T = \{(tr^A_\omega, tr^B_\omega, p_\omega)\}_{\omega\in\Omega}$ is admissible with respect to W. Since $\tau|_{h^t}$ is feasible, $tr^i_\omega = tr^i_{t+1}(\omega^{t+1}, \tau|_{h^t}) \in [0, y^i(\omega)]$. By construction, $p_\omega(\cdot,\cdot) \in W$. Last, suppose agent i deviates in the current period by choosing a transfer $\breve{tr}^i \in [0, y^i(\omega)]$. Consider the strategy $\breve{\tau}_i|_{h^t}$ specifying $\breve{\tau}_i|_{h^t}(\omega_{t+1}, h^0) = \breve{tr}^i$ with $\omega_{t+1} = \omega$ and $\breve{\tau}_i|_{h^t * h^d} = \tau_i|_{h^t * h^d}$ for all $d > 0$ and $h^d \in \mathcal{H}$. By Lemma 1

$$
\begin{aligned}
& (1-\beta)u\big(y^i(\omega_{t+1}) - tr^i_{t+1}(\omega^{t+1}, \tau|_{h^t}) + tr^j_{t+1}(\omega^{t+1}, \tau|_{h^t})\big) \\
+ \;& \beta \sum_{\omega_{t+2}\in\Omega} V^i(\tau|_{h^{t+1}})(\omega_{t+2})\pi(\omega_{t+2}) \\
\geq \;& (1-\beta)u\big(y^i(\omega_{t+1}) - tr^i_{t+1}(\omega^{t+1}, (\breve{\tau}_i|_{h^t}, \tau_j|_{h^t})) + tr^j_{t+1}(\omega^{t+1}, (\breve{\tau}_i|_{h^t}, \tau_j|_{h^t})) \\
+ \;& \beta \sum_{\omega_{t+2}\in\Omega} V^i(\breve{\tau}_i|_{h^t * \breve{h}}, \tau_j|_{h^t * \breve{h}})(\omega_{t+2})\pi(\omega_{t+2})
\end{aligned}
$$

with $\breve{h} = \big(\omega_{t+1}, tr^A_{t+1}(\omega^{t+1}, (\breve{\tau}_A|_{h^t}, \tau_B|_{h^t})), tr^B_{t+1}(\omega^{t+1}, (\breve{\tau}_A|_{h^t}, \tau_B|_{h^t}))\big)$. By definition this is equivalent to

$$
\begin{aligned}
& (1-\beta)u\big(y^i(\omega_{t+1}) - tr^i_{t+1}(\omega^{t+1}, \tau|_{h^t}) + tr^j_{t+1}(\omega^{t+1}, \tau|_{h^t})\big) + \beta V^i(\tau|_{h^{t+1}}) \\
\geq \;& (1-\beta)u\big(y^i(\omega_{t+1}) - tr^i_{t+1}(\omega^{t+1}, (\breve{\tau}_i|_{h^t}, \tau_j|_{h^t})) + tr^j_{t+1}(\omega^{t+1}, (\breve{\tau}_i|_{h^t}, \tau_j|_{h^t})) \\
+ \;& \beta V^i(\breve{\tau}_i|_{h^t * \breve{h}}, \tau_j|_{h^t * \breve{h}})
\end{aligned}
$$

Substituting the above definitions in the last inequality and noting that

$$\begin{aligned} tr^i_{t+1}(\omega^{t+1}, (\breve{\tau}_i|_{h^t}, \tau_j|_{h^t})) &= \breve{\tau}_i|_{h^t}(\omega_{t+1}, h^0) = \breve{tr}^i \\ tr^j_{t+1}(\omega^{t+1}, (\breve{\tau}_i|_{h^t}, \tau_j|_{h^t})) &= \tau_j|_{h^t}(\omega_{t+1}, h^0) = tr^j(\omega^{t+1}, \tau|_{h^t}) = tr^j_\omega \end{aligned}$$

yields

$$\begin{aligned} & (1-\beta)u(y^i(\omega) - tr^i_\omega + tr^j_\omega) + \beta p^i_\omega(tr^i_\omega, tr^j_\omega) \\ \geq\ & (1-\beta)u(y^i(\omega) - \breve{tr}^i + tr^j_\omega) + \beta p^i_\omega(\breve{tr}^i, tr^j_\omega). \end{aligned}$$

Hence, the last condition is also satisfied. The constructed $T = \{(tr^A_\omega, tr^B_\omega, p_\omega)\}_{\omega\in\Omega}$ is admissible with respect to W and supports (w^A, w^B) implying $(w^A, w^B) \in W \subseteq B(W)$. $\square$

Proof of Lemma 4 (By construction). Let $(v^A, v^B) \in L$. Since $L \in B(L)$ there exists a $T = \{(tr^A_\omega, tr^B_\omega, p_\omega)\}_{\omega\in\Omega}$ admissible with respect to L such that

$$v^i = (1-\beta)u(y^i(\omega) - tr^i_\omega + tr^j_\omega) + \beta p^i_\omega(tr^i_\omega, tr^j_\omega)$$

Next, construct a subgame perfect strategy profile supporting (v^A, v^B). The idea here is to unfold the pair (v^A, v^B). First, define first period actions $\tau_i(\omega_1, h^0) = tr^i_{\omega_1}$. Now, construct the actions in period 2. These need to be defined for three cases regarding the first period: A and B complied, A complied but B deviated and A deviated while B complied.

1. Suppose A and B did comply in the first period, i.e. $\widetilde{h}^1 = (\omega_1, tr^A_{\omega_1}, tr^B_{\omega_1})$. This implies A and B receive the pair $p_{\omega_1}(tr^A_{\omega_1}, tr^B_{\omega_1})$. Since L is self-generating, $p_{\omega_1}(tr^A_{\omega_1}, tr^B_{\omega_1}) \in L$. This implies there exists an admissible

$T = \{(tr^A_{\omega_2}, tr^B_{\omega_2}, p_{\omega_2})\}_{\omega_2 \in \Omega}$ with respect to L supporting $p_{\omega_1}(tr^A_{\omega_1}, tr^B_{\omega_1})$. Define $\tau_i(\omega_2, h^1) = tr^i_{\omega_2}$.

2. Suppose i complied but j deviated to $\widetilde{tr}^j \neq tr^j_{\omega_1} \in [0, y^j(\omega_1)]$ in the first period, i.e. $\widetilde{h}^1 = (\omega, \widetilde{tr}^j, tr^i(\omega_1))$. This implies A and B receive the pair $p_{\omega_1}(\widetilde{tr}^j, tr^i_{\omega_1})$. Since L is self-generating, $p_{\omega_1}(\widetilde{tr}^j, tr^i_{\omega_1}) \in L$. This implies there exists an admissible $T = \{(tr^A_{\widetilde{\omega}_2}, tr^B_{\widetilde{\omega}_2}, p_{\widetilde{\omega}_2})\}_{\widetilde{\omega}_2 \in \Omega}$ with respect to L supporting $p_{\omega_1}(\widetilde{tr}^j, tr^i_{\omega_1})$. Define $\tau_i(\widetilde{\omega}_2, \widetilde{h}^1) = tr^i_{\widetilde{\omega}_2}$.

This way one can construct recursively a strategy profile $\tau = (\tau_A, \tau_B)$ supporting $(v^A, v^B) \in L$. After each history the agents' actions form an admissible object T. This means that the agents' actions are feasible and that no agent after any history has an incentive to deviate. By the one-deviation lemma this implies that τ is an subgame perfect strategy profile. □

Proof of Lemma 5 Note first, that $B(L) \in 2^D$ with $D = [u(0), u(1)]^2$. Hence, $B(L)$ is bounded. Consider a sequence $\{(w^A_n, w^B_n)\}_{n=1}^{\infty}$ with $(w^A_n, w^B_n) \in B(L)$ for all n and $\lim_{n\to\infty}(w^A_n, w^B_n) = (w^A, w^B)$ in the Euclidean metric. It remains to show that $(w^A, w^B) \in B(L)$. Every (w^A_n, w^B_n) is supported by some $T_n = \{(tr^A_{n\omega}, tr^B_{n\omega}, p_{n\omega})\}_{\omega \in \Omega}$ which is admissible with respect to L. The function $p_{n\omega}(\cdot)$ is an element of the function space $\{g : [0,1]^2 \to L | L \subseteq R^2 \text{ compact}\}$. This function space is compact in the sup norm. Each of the $3 \times N$ elements of T_n lies in a compact set. Then there exists a convergent subsequence. Let its limit be $T = \{(tr^A_\omega, tr^B_\omega, p_\omega)\}_{\omega \in \Omega}$. Obviously, T supports (w^A, w^B) and since T is admissible with respect to L, $(w^A, w^B) \in B(L)$. □

Proof of Lemma 6 Consider some $(w^A, w^B) \in B\big(hyp(f)\big)$ supported by

$\{(tr^A_\omega, tr^B_\omega, w^A_\omega, w^B_\omega)\}_{\omega\in\Omega}$. Next, define a function on $[0,1]^2$

$$p_\omega(x,y) = \begin{cases} (w^A_\omega, w^B_\omega), & \text{if } (x,y) = (tr^A_\omega, tr^B_\omega) \\ (V_{aut}, V_{aut}), & \text{otherwise.} \end{cases}$$

It remains to show that $\{(tr^A_\omega, tr^B_\omega, p_\omega)\}_{\omega\in\Omega}$ admissible with respect to W. First, by assumption $tr^i_\omega \in [0, y^i(\omega)]$. Second, $p_\omega : [0,1]^2 \to hyp(f)$. Third, by assumption for all i and for all $\hat{tr}^i \in [0, y^i(\omega)]$

$$\begin{aligned} & (1-\beta)u(y^i(\omega) - tr^i_\omega + tr^j_\omega) + \beta w^i_\omega \\ \geq \; & (1-\beta)u(y^i(\omega) - \hat{tr}^i + tr^j_\omega) + \beta V_{aut}. \end{aligned}$$

This yields

$$\begin{aligned} & (1-\beta)u(y^i(\omega) - tr^i_\omega + tr^j_\omega) + \beta p^i_\omega(tr^i_\omega, tr^j_\omega) \\ \geq \; & (1-\beta)u(y^i(\omega) - \hat{tr}^i + tr^j_\omega) + \beta p^i_\omega(\hat{tr}^i, tr^j_\omega). \end{aligned}$$

□

Chapter 2

An Empirical Analysis of the Relationship between Income and Democracy

2.1 Introduction

Identifying the determinants of democracy is of broad interest in economics and political science. Anecdotal evidence suggests that there are rich and democratic countries such as in Western Europe as well as poor and non-democratic countries such as the sub-Saharan countries. Also, in the last decades relatively more countries have become rich and relatively more countries have become democratic. The prime examples are Taiwan and South Korea. This has led scholars to hypothesize that economic development leads to political development or that higher income leads to democracy. The question of interest is then, whether there is empirical evidence supporting this claim.

Measuring income is a straightforward accounting exercise in multiplying and adding prices and quantities. How to measure democracy best is an ongoing debate among researchers however. One group of scholars has proposed to measure democracy by dichotomously viewing democracy as a "question of kind." Another group of scholars views democracy as a continuous concept suggesting a graded scale as a measurement tool. Irrespective of the particular position, there is a broad consensus to measure democracy using an ordinal scale. It is standard for empirical analysis in economics to treat ordered discrete data explicitly as such, by employing an econometric model such as the ordered logit model. Applying standard econometric models in the current context is further complicated by the fact that researchers typically want to control for past democracy levels. This approach requires a way to handle ordinal dependent and ordinal independent variables simultaneously. This paper contributes to the existing literature by developing and estimating a dynamic model where the ordinal variable has the Markov property. My econometric model is similar to the model by Epstein, Bates, Goldstone, Kristensen, and O'Halloran (2006).[1]

There are various suggestions in the literature on how to precisely interpret the statement *a higher income leads to democracy* econometrically. I investigate whether higher income Granger causes democracy. Granger causality is the appropriate econometric causality concept in this context for two reasons. First, the effect of income on democracy is dynamic. The idea here is that as countries become more wealthy over time they transfer more political rights to their citizens. This makes it seem natural to use the standard causality concept in time series analysis. Second, in contrast to stronger microeconometric causality concepts Granger causality is immediately testable without any further assumptions.

[1]After completing this paper, I learned that Epstein, Bates, Goldstone, Kristensen, and O'Halloran (2006) had already developed a model similar to mine.

Mosconi and Seri (2006) have extended the concept of Granger causality to binary processes. Using a similar Markov approach allows me to apply their definition of Granger causality in my econometric model.

Robinson (2006) points out that most work in the empirical literature investigating whether income causes democracy fails to take into account that democracy might have an effect on income. Granger causality in its stronger form is unidirectional. The reason is that any causality notion requires the cause-effect pair to be not symmetric, i.e., if B causes A, then A cannot cause B. In fact, democracy causing income has some intuitive appeal: Political decision makers responsible for inferior economic outcomes can be voted out of office in a democracy while this is not possible in an autocracy. This paper contributes to the literature by testing whether income Granger causes democracy and whether democracy Granger causes income.

The contribution of the seminal work by Nelson and Plosser (1982) is to point out that macroeconomic time series might be unit root processes. I show that GDP per capita is a unit root process for almost all countries. The implication for the econometric analysis is that GDP per capita is an unbounded variable. Failing to take this into account leads to model misspecification and misleading empirical results when investigating the influence of economic development on political development. This paper contributes to the literature by establishing that GDP per capita is a unit root process for almost all countries, by showing how this leads to model misspecification, and by taking this property correctly into account in the empirical analysis.

My empirical analysis suggests the following. First, the effect of income on democracy depends on the current democracy level. Second, higher income Granger causes democracy for full democracies, which means that a higher income helps to stabilize democracies. Third, higher income Granger causes democracy for medium democracies. This does not mean that a higher income would necessarily lift a country from dictatorship to democracy by itself, but rather that income can have a positive effect on regime type where there is already a somewhat favorable environment. Fourth, I find that Granger causality runs from democracy to income, as well. This implies that income and democracy form a bidirectional feedback system which makes it impossible to apply any notion of causality to income and democracy.

The rest of this paper is organized as follows. Section 2.2 reviews the literature. In section 2.3 I establish three stylized facts regarding income and democracy, and discuss relevant properties of the data. Section 2.4 discusses the concept of democracy as a latent variable while section 2.5 discusses the application of the concept of Granger causality in the current context. In section 2.6, I introduce the econometric model and section 2.7 discusses the estimation and the results. In section 2.8 I discuss the evidence regarding the question whether democracy Granger causes income. Section 2.9 summarizes and concludes.

2.2 Discussion of the Literature

The question of whether a higher income leads to a rise in democracy was first empirically investigated by Barro (1999). Barro uses the Freedom House (2009) data for 1972, 1975, 1980, 1985, 1990, 1995 and the data from Bollen (1990) for 1960 and 1965. Bollen's data are composed of various studies and range from 0 (non-democracy) to 100 (democracy). They are however, ordinal. Freedom House assigns seven different values to measure the political rights of a country with 1 being a perfect democracy and 7 being a complete non-democracy. Subsequently, Barro converts all data to a scale from 0 (non-democracy) to 1 (democracy). The econometric model is

$$\text{DEMOC}_{it} = a_{0t} + a_1\text{DEMOC}_{i,t-T} + a_2\text{DEMOC}_{i,t-2T} + a_3\mathbf{Z}_{i,t-T} + u_{it}$$

where i denotes the country, DEMOC is the democracy variable, a_{0t} is a time dummy and $\mathbf{Z}$ contains various measures such as per capita GDP and education. The dependent variable is observed first in 1972. The estimation method is seemingly unrelated regressions. The estimate for past income is 0.058 with a standard error of 0.016.

This econometric analysis leaves some room for improvement. First, assume there are seven democracy categories, 1 (best) through 7 (worst) and assume country A receives score 1 while country B receives score 2. Treating ordinal data as cardinal is equivalent to asserting that country B being less democratic than country A implies that country B is half as democratic as country A. This conclusion is wrong since the ordinal scale of 1 through 7 is invariant to any monotonic transformation while the cardinal scale is not. Treating ordinal variables as cardinal yields unreliable parameter estimates at best.

Second, Barro uses the level of income as the measure of income. As shown in table 3.5 on page 94, the GDP per capita time series for almost all countries have a unit root. This finding has been confirmed in numerous studies since the seminal contribution by Nelson and Plosser (1982). The problem is that a significance test for coefficients in that case are misleading as argued by Granger and Newbold (1974). Granger and Newbold (1974) call the phenomenon of highly significant coefficients in the absence of any economic meaning "spurious regressions."

Third, since GDP per capita is a unit root process, it is necessarily unbounded.

To see this, let y_t denote GDP per capita in period t, assume $\mathrm{E}[u_t] = 0$ and let u_t be serially independent. Then,

$$y_t = y_{t-1} + u_t = \sum_{j=1}^{t} u_j + y_0.$$

Now, letting $a < 0$ and $b > 0$

$$\begin{aligned} \Pr[y_t \in (a,b)] &= \Pr\left[\frac{y_t}{\sqrt{t}} \in \left(\frac{a}{\sqrt{t}}, \frac{b}{\sqrt{t}}\right)\right] \\ &= \Pr\left[\frac{1}{\sqrt{t}}\sum_{j=1}^{t} u_j + \frac{y_0}{\sqrt{t}} \in \left(\frac{a}{\sqrt{t}}, \frac{b}{\sqrt{t}}\right)\right]. \end{aligned}$$

For $t \to \infty$, $\frac{y_0}{\sqrt{t}}$, $\frac{a}{\sqrt{t}}$ and $\frac{b}{\sqrt{t}}$ go to zero while $\frac{1}{\sqrt{t}} \sum_{j=1}^{t} u_j \sim \mathrm{N}(0, \sigma^2)$ by the central limit theorem. This implies that

$$\Pr[y_t \in (a,b)] = 0$$

for lim $t \to \infty$ or in other words, that y_t is unbounded. Then, for large t

$$a_{0t} + a_1 \mathrm{DEMOC}_{i,t-T} + a_2 \mathrm{DEMOC}_{i,t-2T} + a_3 \mathbf{Z}_{i,t-T} + u_{it} \notin [0,1]$$

contradicting the fact that $\mathrm{DEMOC}_{it} \in [0,1]$ by construction of the new democracy variable. In other words,

$$\begin{aligned} &\Pr\Big[\mathrm{E}\big[\mathrm{DEMOC}_{it} \mid \mathrm{DEMOC}_{i,t-T}, \mathrm{DEMOC}_{i,t-2T}, \mathbf{Z}_{i,t-T}\big] \\ = \quad & a_{0t} + a_1 \mathrm{DEMOC}_{i,t-T} + a_2 \mathrm{DEMOC}_{i,t-2T} + a_3 \mathbf{Z}_{i,t-T}\Big] < 1 \end{aligned}$$

implying that the model is not a conditional expectation model. More generally, many econometric models can be written as

$$Y = f(X, \beta, U) \tag{2.1}$$

with Y, X and U having realizations in $\mathcal{S}_y, \mathcal{S}_x$ and $\mathcal{S}_u$, respectively. A fundamental econometric principle is that first, for all $\hat{y} \in \mathcal{S}_y$ there exists at least one $\hat{x} \in \mathcal{S}_x$ and $\hat{u} \in \mathcal{S}_u$ such that (1) holds and second, for all $\tilde{x} \in \mathcal{S}_x$ and $\tilde{u} \in \mathcal{S}_u$ there exists one and only one $\tilde{y} \in \mathcal{S}_y$ such that (1) holds. Bierens (2009) calls this the feasibility rule. A violation of the feasibility rule means that the model in (1) is only valid for proper subspaces of $\mathcal{S}_y, \mathcal{S}_x$ and $\mathcal{S}_u$. This is one case of model misspecification. Barro (1999) notes himself that in his analysis the most negative fitted value is -0.02 and the highest is 1.09. This makes it clear that Barro's approach suffers from model misspecification.[2]

Two more issues are noteworthy. First, there is a censoring issue. Most OECD countries have been democracies since World War II. Also, they have experienced substantial economic growth. Within the data however, these look like a group of countries where significant economic development had no impact on regime type resulting in a downward bias. Second, there is the issue of unexploited data. That is, there is a lot of information left on the table since the authors use 5-year intervals. The problem is that we do not know much about the hypothesized

[2] One way to circumvent this problem is to allow for heteroskedasticity. The idea is that

$$0 \leq a_{0t} + a_1 \mathrm{DEMOC}_{i,t-T} + a_2 \mathrm{DEMOC}_{i,t-2T} + a_3 \mathbf{Z}_{i,t-T} + u_{it} \leq 1$$

or

$$\begin{aligned} & -a_{0t} - a_1 \mathrm{DEMOC}_{i,t-T} - a_2 \mathrm{DEMOC}_{i,t-2T} - a_3 \mathbf{Z}_{i,t-T} \\ \leq\ & u_{it} \\ \leq\ & 1 - a_{0t} - a_1 \mathrm{DEMOC}_{i,t-T} - a_2 \mathrm{DEMOC}_{i,t-2T} - a_3 \mathbf{Z}_{i,t-T}. \end{aligned}$$

This however, conflicts with the conditional mean independence assumption.

effect's time horizon. For this reason it is desirable to exploit as much information as possible.

Acemoglu, Johnson, Robinson, and Yared (2008) try to improve upon Barro's analysis on two fronts.[3] First, they are concerned with unobserved heterogeneity across countries which leads them to include country fixed effects. Second, they would like to know whether higher income causes democracy. In order to interpret the regression results in terms of causation they want to make sure that the conditional mean independence assumption holds. This leads them to use past savings and trade-weighted world income as instrumental variables for income. This however, does not solve the two main issues. That is, the data are ordered and discrete and even if a transformation to a $[0, 1]$ scale would be valid, the econometric model must make sure that the right-hand side is bounded between 0 and 1.

The relationship between income and democracy has also received attention in the political science literature. The seminal contribution is the investigation by Lipset (1959). The notion that income causes democracy is known as 'Lipset hypothesis' or as integral part of the modernization theory. There is no consensus in the empirical literature. Przeworski, Alvarez, Cheibub, and Limongi (2000) find no evidence for the modernization theory while Epstein, Bates, Goldstone, Kristensen, and O'Halloran (2006) find evidence supporting the claim that income causes democracy. Both studies account for the fact that the democracy measure is discrete and ordered. They fail however, to account for the fact that GDP per capita is a unit root process leading to spurious transition probabilities. Also, as Robinson (2006) points out, it is unclear in these studies what econometric causality concept they employ. The next section introduces the data.

[3]For a critique of Acemoglu, Johnson, Robinson, and Yared (2008) see Gundlach and Paldam (2008).

2.3 Data and Stylized Facts

This section has three parts. First, I present the data. Then, I document three stylized facts in the data. Finally, I discuss some properties of the data.

2.3.1 Data

The data on real GDP per capita are from Heston, Summers, and Aten (2006).[4] The data on democracy from Freedom House (2009) is my main source. I also use the data on political rights. These are ordinal, ranking from 1 (best) to 7 (worst). Data are available for 192 countries from 1973 to 2008. The panel however, is unbalanced. The scores for a given country in a given year are based on three categories: electoral process, pluralism and participation, and the functioning of the government. I also use the widely used Polity IV data set from Marshall and Jaggers (2004). The main measure of democracy is the polity index which ranges from -10 to 10. It is obtained by subtracting the autocracy index ($0-10$) from the democracy index ($0-10$). The democracy index captures political participation, constraints on the executive and civil liberties. The autocracy index reflects the openness and competitiveness of executive recruitment, the constraints on the executive and as well as the competitiveness of political participation. For the spatial analysis I use the data by Gleditsch and Ward (2001) to determine a country's neighboring countries in 1973 and 2002.

2.3.2 Three Stylized Facts

Consider table 2.1. The third column gives the number of very democratic countries, the fifth column the number of very non-democratic countries and the fourth

[4]The data (rgdpch) are chain$-$weighted and given in international dollars of 2000.

column gives the number of countries in between. The second row displays the countries above the 80th percentile of the log GDP per capita distribution in 1973. Most of the rich countries are democratic. In contrast, the third row displays the countries below the 20th percentile of the income distribution in 1973. Most of the poor countries are non-democratic. The fifth and sixth row confirm this pattern for 2003. So, the first stylized fact is that rich countries tend to be democratic while poor countries tend to be non-democratic.

	Year	$\{1,2\}$	$\{3,4,5\}$	$\{6,7\}$
all countries	1973	37	34	57
top 20% GDP	1973	19	1	6
bottom 20% GDP	1973	2	7	17
all countries	2003	55	38	35
top 20% GDP	2003	21	2	3
bottom 20% GDP	2003	1	14	11

Table 2.1. Distribution of Political Rights.

The second stylized fact is that world income is increasing and democracy is on the rise. More exactly, the average country has become richer and the distribution is skewed towards higher income. Analogously, the median country has become more democratic and the distribution is skewed towards more political rights. Figure 2.1 on page 31 depicts the distribution of world income in 1973 and 2002. The picture suggest that the left mode carries more mass in 1973 while the middle and the right mode move to the right from 1973 to 2003. Table 2.2 on page 31 quantifies this. The average country has grown from a real log GDP per capita of 8.17 in 1973 to 8.54 in 2003. More pronounced however, is the third moment, i.e.

the skewness. While the distribution was slightly skewed to the left in 1973 (0.34), it is more skewed to the right in 2003 (-0.31). Interestingly, this movement of the distribution and within the distribution seems to be very constant from 1973 to 2003.

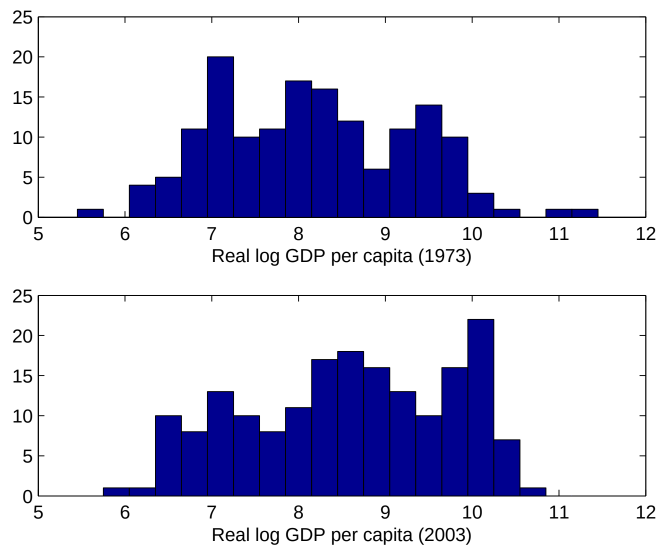


Figure 2.1. Distribution of World Income across Countries in 1973 and 2003.

	1973	1983	1993	2003
mean(log GDP_t^i)	8.17	8.27	8.35	8.54
median(log GDP_t^i)	8.13	8.26	8.42	8.62
skewness(log GDP_t^i)	0.34	0.06	-0.2	-0.31
Countries	130	148	174	177

Table 2.2. Distribution of World Income across Countries.

The dynamics are stronger for political rights. The left picture in figure 2.2 on page 32 shows that there were relatively more dictatorships than democracies in 1973, while this is almost completely reversed in the right picture in 2003. Table

2.1 on page 30 confirms this. According to the first row, there were 37 democracies and 57 non-democracies in 1973. The picture is reversed in 2003. Now, there are 55 democracies and 35 non-democracies. This implies that there are a large number of countries that have made the transition from autocracy to democracy. Huntington (1991) calls this the *Third Wave of Democratization*.

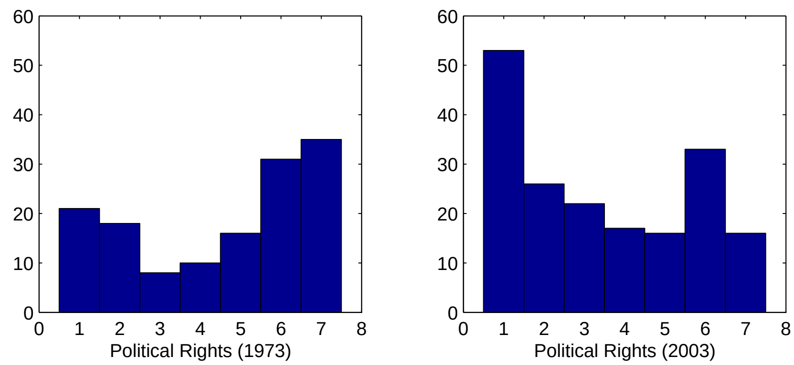


Figure 2.2. Distribution of Political Rights across Countries in 1973 and 2003.

The third stylized fact is that neighboring countries have similar income and regime types. Figure 2.4 on page 51 is the "Map of Freedom" by Freedom House (2008). Dark (blue) countries are not free, light (yellow) are partly free and the remaining countries are free.[5] This map illustrates the stylized fact for the case of regime types. For example, countries in Europe and North America are free, while countries in Northern Africa and the Middle East are mostly non-democratic. Countries in Latin America and Southeastern Africa however, are partly free. Table 2.3 quantifies this effect.[6] This table shows that a country's regime type and and income are significantly correlated with its neighbors' regime type and income, respectively. Furthermore, this correlation seems to be time-invariant.

[5]The Freedom House measures "Free", "Partly Free" and "Not Free" are given by a combination of the measures of political rights and civil liberties.

[6]Neighboring countries are defined as having zero minimum distance. Also, in order to facilitate the presentation I calculate correlation coefficients between ordinal political rights data

	1973	2002
$\text{Corr}\left(\text{PR}_t^i, \text{PR}_t^{\text{NBs(i)}}\right)$	0.53	0.68
Countries	117	147
$\text{Corr}\left(\log \text{GDP}_t^i, \log \text{GDP}_t^{\text{NBs(i)}}\right)$	0.8	0.78
Countries	80	129

Table 2.3. Spatial Correlation between Political Rights and Income.

This fact is not surprising if one assumes that neighboring countries are similar with respect to certain fundamentals. However, it is surprising if one assumes that regime type and income are equilibrium outcomes. Alesina and Spolaore (1997) for example, determine the number and size of countries endogenously. The idea is that there is a tradeoff between heterogeneity and governability. From this perspective, it is not obvious why neighboring countries should be similar.

2.3.3 Properties of the Data

The data on democracy and income have several relevant properties. First, the data on democracy is fairly persistent. Table 2.4 reports the number of transitions from a democracy score today to a democracy score tomorrow. Most of the observations are located around the diagonal axis indicating that usually regime type does not change dramatically from year to year. Figure 2.3 on page 35 depicts this fact graphically. Second, log GDP per capita is a unit root process in almost all countries. Table 2.14 on page 59 shows that only for a few countries the null hypothesis that a unit root process is present can be rejected. The important implication is that regressing democracy on income might lead to spurious results.

while incorrectly assuming these data are cardinal.

	1	2	3	4	5	6	7
1	818	16	2	0	0	0	1
2	30	450	34	5	1	7	2
3	2	35	184	30	6	3	2
4	0	15	35	229	30	9	7
5	0	10	7	40	384	48	4
6	1	5	5	17	65	603	52
7	0	2	1	5	4	73	561

Table 2.4. Transitions from Democracy Today to Tomorrow.

Third, there is a lot of heterogeneity in the data. In order to get a better understanding of the amount of heterogeneity in the data I do the following. First, for each country I count the number of changes in the political rights score for the better and the number of changes for the worse between 1973 and 2003. Then, I find the countries that have more positive changes than negative changes and vice versa. The latter countries can be found in the first row of Table 2.5 while the former countries can be found in the second row. Then, for all of these countries I determine whether they have grown below or above the median between 1973 and 2003. The numerical difference between the first and the second row reflects the fact that the world has become more democratic in the last decades. The fact that the second and the third column are fairly similar seems to suggest that the direction of a country's political development is not correlated with its relative economic growth rate. In particular, there seems to be a significant group where (relative) economic and political development move into opposite directions. The next section discusses the concept of democracy as a latent variable.

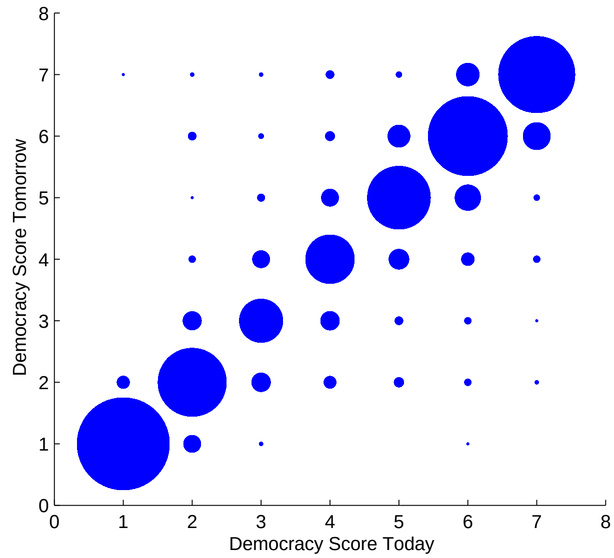


Figure 2.3. Transitions from Democracy Today to Democracy Tomorrow.

	above median growth	below median growth
negative net change in PR	12	9
positive net change in PR	28	33
zero net change in PR	24	22

Table 2.5. Countries with Political Change and Economic Growth.

2.4 Democracy as a Latent Variable

In order to avoid the problems discussed in section 2.2, I treat democracy explicitly as an ordered discrete variable. This is achieved by modeling democracy as a latent unobservable variable.

There is a debate in political science regarding how to measure democracy. Among others, Alvarez, Cheibub, Limongi, and Przeworski (1996) have proposed to measure democracy dichotomously. The idea is that democracy "is a question of kind before it is a question of degree" [Elkins (2000)]. Putting it differently, political systems ought of be thought of as "bounded wholes." Also, according to Alvarez, Cheibub, Limongi, and Przeworski (1996) measuring democracy dichotomously reduces measurement error. In contrast, Bollen and Jackman (1989) view democracy as a continuous concept implying that a graded scale is the superior measurement tool. Obviously, a graded scale yields more variation to be exploited than a dichotomous scale. While proponents of a dichotomous scale refer to Huntington (1991) for a justification in political theory, proponents of a graded scale refer to Dahl (1971). According to Dahl (1971) various regimes display various degrees of democratization whereby these various degrees are relative with respect to a theoretical limit. This theoretical limit of a perfect democracy is similar to the theoretical construct of the perfect vacuum. Here, I adopt this notion of a latent continuous democracy variable.[7] While I view the positive end of the continuum as a theoretical construct of a perfect democracy I do not take a stand as to whether I view the negative end as a stable dictatorship or absolute anarchy. Rather, I view this negative end as absolute non-democracy. Also, my continuous underlying latent variable is unidimensional which reflects a simplification rather than a deliberate choice against a multidimensional latent democracy variable.

2.5 Granger Causality

In this section I discuss the concept of Granger causality and explain why I prefer to use it in the current context. Granger (1969) explains his concept as follows.

> We say that Y_t is causing X_t if we are better able to predict X_t using all

[7] Treier and Jackman (2008) model democracy explicitly as a latent variable in order to assess the measurement error in the Polity IV data set.

available information than if the information apart from Y_t had been used.

The concept of Granger causality has two distinctive features. First, Granger wants to exploit the fundamental notion that the cause must precede the effect. In other words, the temporal ordering of events is irreversible. This implies that Granger causality is a natural concept in a dynamic setting. Second, Granger's definition contains explicitly a criterion for causality which is immediately applicable. According to the above definition, Y is Granger causing X if past values of Y help us to predict X given past values of X. So, the great advantage of this concept is that its causality notion is immediately testable without further assumptions.

The fact that predictability is used as criterion for causality has attracted some criticism. This criticism claims that the intuitive notion of causality is the notion of controlled variation. In other words, Y causes X if, holding all relevant factors constant, a change in Y leads to a change in X. For example, Leamer (1985) points out that the term "causality" is strongly misleading here since Granger's definition actually tests for precedence. His point is then the following: "Weather forecasts regularly precede the weather, but few of us take this as evidence that the forecasts 'cause' the weather." Leamer is absolutely right. However, in this example weather would also Granger cause weather forecasts because weather forecasts obviously use past weather data. So, the empirical analysis based on the concept of Granger causality would find a bidirectional feedback system in Leamer's example.[8] Finding a bidirectional feedback system in which X Granger causes Y and Y Granger causes X is of limited empirical interest as it indicates that X and Y are endogenous in the same system. The idea of Sims (1972) on the other hand is, that if Y Granger

[8] For an interesting discussion, see Jaeger and Paserman (2008). They apply the concept of Granger causality to the Palestinian-Israeli Conflict.

causes X while X fails to Granger causes Y we must have that Y is exogenous relative to X. This is the strong sense of unidirectional Granger causality. Presence of Granger causality in one direction in conjunction with the absence of Granger causality in the other direction restricts the set of models explaining this pattern considerably as any endogeneity is ruled out.

Hence, my stand is this: Granger causality is weaker than the causality notion based on controlled variation. However, unidirectional Granger causality, i.e., Y is Granger causing X, but X is not Granger causing Y is stronger than a simple statistical data description.[9]

Two well-known issues with the concept of Granger causality deserve attention. First, Leamer (1985) stresses that the Granger causality estimates are highly sensitive to the composition of the information set. In particular, Sims (1972) finds that money Granger causes but controlling for interest rates leads Sims (1980) to conclude that income Granger causes money. Second, the concept of Granger causality is based on the notion that the temporal ordering of events is irreversible. This however, might be challenged once expectations are introduced. One possibility is that the expectation of a future rise in democracy is associated with the expectation of stronger property rights, leading to more investment today. Even though this is a clear possibility, I assume that there are no relevant expectations involved here.

Next, I give some explanations as to why I prefer applying Granger causality in the current context. First, Granger causality is the standard causality concept in time series analysis and we are interested in a dynamic effect. The idea is

[9] Stock and Watson (2001) argue for Granger causality to be a useful tool for data description. Attanasio, Picci, and Scorcu (2000) have recently used the concept of Granger causality to provide an exhaustive empirical analysis of the relationships between savings, investment and growth using various panel data sets.

that some countries get wealthier over time and subsequently transfer more and more political rights to its citizens. Typical recent examples are South Korea or Taiwan. Even though one might argue that in principle all causality phenomena are dynamic, it seems that the current phenomenon is somewhat more dynamic in nature than job training programs. Secondly, Granger causality is testable without any additional assumptions. This leads me to conclude that Granger causality is the appropriate econometric causality concept in this context.[10]

Granger causality is typically applied to observable data. Here however, latent (unobserved) democracy is the variable of interest. Mosconi and Seri (2006) develop a general framework for bivariate binary time series. In their framework they define Granger causality rigorously. The authors make the concept of Granger causality operational by assuming that the binary process is a first-order Markov process. The idea is then to use the stationary transition probabilities for an operational definition of Granger causality. By making a similar Markov assumption in my econometric model their definitions apply here.

The next section introduces the econometric model. The main difficulty is to deal with serial dependence in the ordered discrete variable.[11]

2.6 Econometric Model

There are T periods and N countries. For each country i we observe in each period t an ordered discrete democracy variable d_{it} and GDP per capita y_{it}. Since y_{it} is a

[10]Hurlin and Venet (2001) propose a particular way how to apply Granger causality to panel data.

[11]My model is similar to the model of Epstein, Bates, Goldstone, Kristensen, and O'Halloran (2006).

unit root process for almost all countries, I take first differences

$$\triangle y_{it} = y_{it} - y_{it-1}.$$

From here onwards, I use only country i's log growth in period t, $\triangle y_{it}$. The observations are i.i.d across countries. There are J ordered democracy categories $\{1, ..., J\}$. The observed democracy category d_{it} is determined by the true latent democracy variable d_{it}^*. The objective is to investigate whether income Granger causes latent democracy. This suggests to regress latent current democracy on latent past democracy and past income. The difficulty is in dealing with latent variables on both sides of the equation simultaneously.[12] In order to deal with this problem, I make the following assumption.

Assumption 1 (Markov) *For all i and t,*

$$\mathrm{P}(d_{it}|d_{it-1}, ..., d_{i0}, \triangle y_{iT}, ..., \triangle y_{i0}) = \mathrm{P}(d_{it}|d_{it-1}, \triangle y_{it-1}, ..., \triangle y_{it-S}).$$

This assumption allows for serial dependence in the ordinal variable. In particular, assume country i's democracy score in year $t-1$ is j, i.e. $d_{it-1} = j \in \{1, ..., J\}$. Then,

$$d_{it} = h \Leftrightarrow k_{h-1,j} < d_{it}^* \leq k_{h,j}$$

for $h = 1, ..., J$ with $k_{0,j} = -\infty, k_{J,j} = +\infty$ and $k_{h-1,j} < k_{h,j}$. The question of interest is whether past income helps to explain latent current democracy conditional

[12] See Ronning and Kukuk (1996) for a model with latent variables being dependent and independent variables.

on latent past democracy, so

$$d_{it}^{*} = \alpha_j + \sum_{s=1}^{S} \beta_{js} \triangle y_{it-s} + u_{it}$$

with u_{it} being distributed according to the logistic distribution with mean 0 and variance $\frac{\pi^2}{3}$. Note that I do not include country fixed effects. The reason is that $\{y_{it}\}$ is a unit root process. In order to avoid spurious results I take first differences which cancels out all fixed effects or initial values. Also, I do not include any control variables. Granger's causality concept prescribes the inclusion of all "relevant information". To my knowledge however, there is little consensus in political theory on the predictors of democracy. Then,

$$d_{it}^{*} | d_{it-1} = j, \triangle y_{it-1}, ..., \triangle y_{it-S}$$

is distributed according to the logistic distribution with mean

$$\alpha_j + \sum_{s=1}^{S} \beta_{js} \triangle y_{it-s}$$

and variance $\frac{\pi^2}{3}$. This in turn yields the conditional probability of d_{it}

$$\begin{aligned} & \mathrm{P}(d_{it} = h | d_{it-1} = j, \triangle y_{it-1}, ..., \triangle y_{it-S}) \\ = & \int_{k_{h-1,j}}^{k_{h,j}} \frac{\exp(-d_{it}^{*} + \alpha_j + \sum_{s=1}^{S} \beta_{js} \triangle y_{it-s})}{(1 + \exp(-d_{it}^{*} + \alpha_j + \sum_{s=1}^{S} \beta_{js} \triangle y_{it-s}))^2} \, \mathrm{d}d_{it}^{*} \end{aligned}$$

with $\mathrm{P}(d_{it} = J | d_{it-1} = j, \triangle y_{it-1}, ..., \triangle y_{it-S}) = 1 - \sum_{h=1}^{J-1} \mathrm{P}(d_{it} = h | d_{it-1} = j, \triangle y_{it-1}, ..., \triangle y_{it-S})$.

For each country i I am interested in maximizing the conditional probability

$$\mathrm{P}(d_{iT}, ..., d_{i1} | d_{i0}, \triangle y_{iT}, ..., \triangle y_{i0}).$$

Since the data for all countries are drawn independently, I am interested in maximizing

$$\prod_{i=1}^{N} \mathrm{P}(d_{iT}, ..., d_{i1} | d_{i0}, \triangle y_{iT}, ..., \triangle y_{i0}).$$

Applying the definition of the conditional distribution repeatedly yields

$$\mathrm{P}(d_{iT}, ..., d_{i1} | d_{i0}, \triangle y_{iT}, ..., \triangle y_{i0}) = \prod_{t=1}^{T} \mathrm{P}(d_{it} | d_{it-1}, ..., d_{i0}, \triangle y_{iT}, ..., \triangle y_{i0}).$$

Using the Markov assumption above, I have

$$\mathrm{P}(d_{iT}, ..., d_{i1} | d_{i0}, \triangle y_{iT}, ..., \triangle y_{i0}) = \prod_{t=1}^{T} \mathrm{P}(d_{it} | d_{it-1}, \triangle y_{it-1}, ..., \triangle y_{it-S}).$$

Collecting all parameters into the vector θ yields

$$\theta = \{\alpha_j, \beta_{j1}, ..., \beta_{jS}, k_{1,j}, ..., k_{J-1,j}\}_{j=1}^{J}.$$

Then, $\mathrm{P}(d_{it} | d_{it-1}, \triangle y_{it-1}, ..., \triangle y_{it-S}; \theta)$

$$= \sum_{h=1}^{J} \sum_{j=1}^{J} \mathbf{1}(d_{it} = h, d_{it-1} = j) \mathrm{P}(d_{it} = h | d_{it-1} = j, \triangle y_{it-1}, ..., \triangle y_{it-S}).$$

The log-likelihood function can be written as follows:

$$L(\theta) \quad = \quad \log \prod_{i=1}^{N} \prod_{t=1}^{T} \mathrm{P}(d_{it} | d_{it-1}, \triangle y_{it-1}, ..., \triangle y_{it-S}; \theta)$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T} \log \mathrm{P}(d_{it}|d_{it-1}, \triangle y_{it-1}, ..., \triangle y_{it-S}; \theta)$$
$$= \sum_{i=1}^{N}\sum_{t=1}^{T} \log \mathrm{P}_{it}(\theta).$$

The asymptotic variance is given by the familiar expression

$$\widehat{AVAR}(\hat{\theta}) = \left(\frac{1}{N \cdot T} \sum_{i=1}^{N}\sum_{t=1}^{T} \frac{\partial \log \mathrm{P}_{it}(\hat{\theta})}{\partial \theta} \frac{\partial \log \mathrm{P}_{it}(\hat{\theta})}{\partial \theta'} \right)^{-1}.$$

For the same reasons as in the case of a standard logit model, the identification restrictions are given by

$$\alpha_j = 0$$

for $j = 1, ..., J$. The question of interest is whether the $(\beta_{j1}, ..., \beta_{jS})$ are significantly different from zero. That is, for all $j \in \{1, ..., J\}$,

$$\mathrm{H}_j : \beta_{j1} = ... = \beta_{jS} = 0.$$

The next section discusses the estimation and presents the results.

2.7 Estimation and Results

Instead of estimating the likelihood function directly, I run an ordered logit separately for every democracy level. Each country i in period t has an unique democracy score d_{it-1}. This allows me to partition the data set according to the past democracy score in J disjoint subsets. Data subset $j \in \{1, ..., J\}$ is then used to estimate the parameters $\{\beta_{j1}, ..., \beta_{jS}, k_{1,j}, ..., k_{J-1,j}\}$.

First, I estimate the model using the original Freedomhouse data. The first results are reported in table 2.6 on page 52. The first row presents the estimates of an ordered logit conditional on the democracy in period $(t-1)$ being 1 (perfect). In order to avoid spurious results, I have replaced levels of income with differences in income. A quick glance a the coefficient estimates reveals that they differ substantially across democracy levels. This is an important first lesson: The impact of income on democracy is very different depending on the current level of democracy. The last column presents the results of an F-test with the null that past economic growth is irrelevant. The p-values for past democracy levels of 1 and 4 show that economic growth is significant at the 10% and 6% level, respectively. The natural concern here is that the democracy scale is too refined, so that all significance is crushed.

For these reasons I regroup the data to see whether the results are similar at a more aggregate level. This also makes sense from a conceptual perspective because we are not so much interested in the effect of income on the legislature of campaign contributions but rather whether there are big shifts in the level of political rights. Table 2.7 on page 53 reports the results once the categories are regrouped such that

$$\{1,2\} = 1, \{3,4,5\} = 2, \{6,7\} = 3.$$

The picture is much clearer now. The p-values of the F-test show that past economic growth is highly significant if a country is democratic or medium democratic. Past income however, is only significant at the 20% level if a country is non-democratic. This suggests that for income to have an effect on democracy there needs to be a favorable environment. Putting it differently, income alone does not seem to lift a country from dictatorship to democracy.

The regrouping was somewhat arbitrarily imposed. Therefore, I check the results using Freedomhouse's own three level scale:

$$1 = \textit{Free}, 2 = \textit{Partly Free}, 3 = \textit{Not Free}.$$

Freedomhouse arrives at this scale by averaging political rights and civil liberties. Table 2.8 shows the results. The p-value of the F-test reveals that past economic growth has an effect only if the past democracy level is 1. This is not surprising since there is little theoretical work suggesting that economic growth has an impact on civil liberties. One way to interpret this finding is that past income has only a stabilizing effect on a society's freedom.

So far, I have only included the time period $1973-2003$. This ignores many events in Africa in the 1950s and 1960s where a lot of countries became independent and later turned into dictatorships as well as regime changes in South America during that period. For these reasons, I turn to the widely used Polity IV data set.[13] The main advantage is that I can now cover the time period $1950-2006$. The drawback is that the sample shrinks to 43 countries.[14] The main measure for democracy is the polity index raging from -10 to 10. Following Valentino, Huth, and Balch-Lindsay (2004) I regroup the 21 categories as follows:

$$\{7, ..., 10\} = 1, \{-6, ..., 6\} = 2, \{-10, ..., -7\} = 3.$$

The results are shown in table 2.9 on page 54. The p-values of the F-test show that conditional on a country being very democratic or being very non-democratic past economic growth has a significant effect on democracy. In contrast to the

[13] See Marshall and Jaggers (2004).

[14] ARG, AUL, AUS, BEL, BOL, BRA, CAN, COL, COS, DEN, EGY, ETH, SAL, FIN, FRN, GUA, HON, IND, IRE, ISR, ITA, JPN, MEX, NTH, NEW, NIC, NOR, PAK, PAN, PER, PHI, POR, SAF, SPN, SRI, SWD, SWI, THI, TUR, UKG, USA, URU, VEN.

Freedomhouse data there is an effect for the very non-democratic countries but not for the medium democratic countries. One possible explanation is that there are too many countries bundled in the middle. The other explanation is that adding the 1950s and 1960s adds more drastic regime changes via coup d'etats.

I conclude that the evidence suggests that higher income Granger causes democracy. Income has the strongest effect in democratic countries suggesting that income has a stabilizing effect on democracy.[15] Income seems to have a somewhat weaker significant effect on democracy in medium democratic countries suggesting that income alone does not lift a country from dictatorship to democracy. There are two disclaimers, however. First, I do not control for heterogeneity. Second, the results seem to be sensitive to the way the democracy data are grouped.

2.8 Does Democracy Granger cause Income?

Given the evidence that income Granger causes democracy, it is natural to ask whether the reverse holds. The reason is that any notion of causality is unidirectional. Next, I derive the model used to investigate whether democracy Granger causes income. Note first that $\{y_{it}\}$ is a unit root process with a drift:

$$y_{it} = y_{it-1} + \eta + v_{it},$$

where v_{it} is stationary. Then backwards substitution yields

$$y_{it} = y_{i0} + \eta t + \sum_{j=1}^{t} v_{ij}.$$

[15] The stabilizing effect of income on regime type is also discussed by Lipset (1959).

The initial value y_{i0} captures here the country-specific fixed or random effects. Now suppose that the stationary error term v_{it} is a linear process of present and past values of some variable w_{it}, i.e.,

$$v_{it} = \sum_{j=0}^{\infty} c_j L^j w_{it} = C(L) w_{it},$$

where L is the lag operator. Assume that

Assumption 2 *The lag operator $C(L)$ is invertible, with inverse $C^{-1}(L) = \beta(L) = 1 - \sum_{s=1}^{S} \beta_s L^s$.*

Moreover, suppose that

Assumption 3 $w_{it} = \sum_{k=1}^{K} \delta_k D_{it-1k} + u_{it}$ *with* $\mathrm{E}[u_{it}] = 0, \mathrm{E}[u_{it}^2] = \sigma^2$ *and* u_{it} *distributed independently across countries and periods, where*

$$D_{it-1k} = \begin{cases} 1, & d_{it-1} = k \\ 0, & \textit{otherwise.} \end{cases}$$

Then we have

$$\begin{aligned} \triangle y_{it} &= \eta + v_{it} \\ &= \eta + C(L) w_{it} \\ &= \eta + \beta^{-1}(L) \left[\sum_{k=1}^{K} \delta_k D_{it-1k} + u_{it} \right]. \end{aligned}$$

Multiplying both sides by $\beta(L)$ yields

$$\triangle y_{it} - \sum_{s=1}^{S} \beta_s \triangle y_{it-s} = \beta(1)\eta + \sum_{k=1}^{K} \delta_k D_{it-1k} + u_{it}.$$

Defining $\alpha \equiv \beta(1)\eta$ yields the econometric model

$$\triangle y_{it} = \alpha + \sum_{s=1}^{S} \beta_s \triangle y_{it-s} + \sum_{k=1}^{K} \delta_k D_{it-1k} + u_{it}.$$

Since this is a system of T equations with the restriction that $\beta_t = \beta$ for $t = 1, ..., T$ I have a pooled panel data model that can be estimated by OLS.[16] The results are given in tables 2.10 to 2.13 starting on page 55. The result is that democracy is significant in all specifications for the Freedomhouse data but insignificant for the Polity IV data.

The econometric model is applied to the same four data sets as in the previous section. For each data set I report the specification with one, three and five lagged growth levels. Tables 2.10, 2.11 and 2.12 report the results for the Freedomhouse data with seven democracy categories, three categories regrouped by me and three categories given by Freedomhouse, respectively. The results are very similar across these specifications. First, the Schwarz criterion is the lowest for three lagged dependent variables which implies that this is the best model specification. Second, the p-value of the F-test shows that past democracy levels are highly significant in explaining current growth. Table 2.13 on page 58 reports the results for the Polity IV data set for 43 countries between 1950 and 2006. The Schwarz criterion is the lowest for five lagged dependent variables implying that this is the best specification to use. The p-value of the F-test is 0.91 which means that past democracy is highly insignificant. However, since the sample includes only 43 countries I have not much confidence in this result. Also, the Polity IV data set is artificially smoothed casting some doubts on its reliability. I conclude that past democracy is significant in explaining current growth meaning that democracy

[16] I use EViews to carry out the estimation. The panel options are all the default options.

Granger causes income.

2.9 Interpretation and Conclusion

This paper investigates empirically the relationship between income and democracy. Its main contribution is to develop and estimate a model where the ordered and discrete democracy data have the Markov property, account for the fact that GDP per capita is a unit root process and test whether income causes democracy as well as whether democracy causes income. The purpose of this section is to summarize and interpret the results.

Even though I am interested in the effect of income on democracy, I use economic growth instead of income levels in the empirical analysis. This procedure is inevitable since the GDP per capita time series for almost all countries exhibits a unit root as shown in table 2.14 on page 59. So, strictly speaking my empirical results relate economic growth or changes in income to democracy. Given that the question is whether economic development leads to political development the effect of changes in income on democracy might actually be more interesting from a policy perspective.

Second, as documented in section 2.3 there is a lot of heterogeneity in the data. I do not not include covariates so far since there is little guidance from political theory as to which factors determine the "relevant information." I do however, account for heterogeneity with respect to last period's democracy level. In particular, there is slope heterogeneity with respect to last period's democracy level. This flexible specification accounts for some heterogeneity.

Then, there is the issue of robustness. The parameter estimates seem sensitive to the grouping of the democracy data, the source of the democracy data and the relevant time period. There are at least two issues here. First, there is – as mentioned above – model uncertainty in that there is no consensus in political theory regarding fundamental determinants of democracy. Second, democracy is a difficult concept to measure. This yields different results depending on the data set employed.[17] My position is as follows: I prefer the Freedomhouse data over the Polity IV data because the latter uses artificially smoothed data. Also, Treier and Jackman (2008) raise some issues regarding the Polity IV data set. Then, I prefer a 3 level scale to a 7 or 21 level scale in order to capture broad changes in democracy.

I conclude by summarizing the main empirical findings. First, the effect of income on democracy depends on the current democracy level. Second, higher income Granger causes democracy for full democracies. This means that a high income helps to stabilize democracies. Third, higher income Granger causes democracy for medium democracies. This does not mean that a higher income can lift a country from dictatorship to democracy, but rather that for income to have a positive effect on regime type there needs to be a favorable environment. Fourth, I find that Granger causality runs from democracy to income, as well. This implies that income and democracy form a bidirectional feedback system which makes it impossible to apply any notion of causality to income and democracy.

[17] See Casper and Tufis (2003) for a discussion of various data sets.

2.10 Appendix B

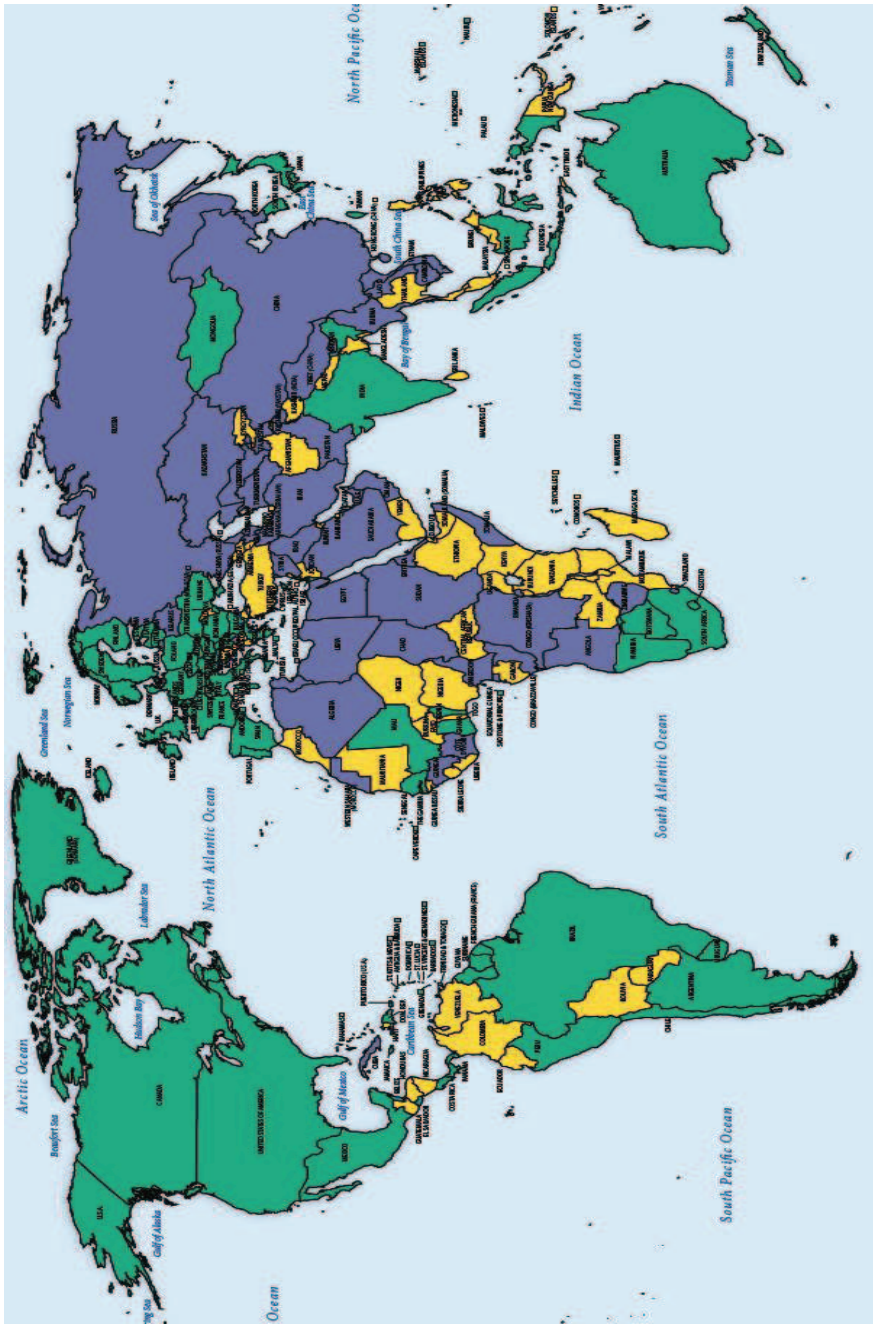


Figure 2.4. Map of the World 2008 by Freedom House.

Democracy Score in $(t-1)$	Growth in $(t-1)$	Growth in $(t-2)$	Growth in $(t-3)$	Growth in $(t-4)$	Growth in $(t-5)$	F-Test
1	14.14	-12.47				2.27 [0.1]
	(1.9)[0.06]	(-1.68)[0.09]				
2	-4.18	-2.07				1.96 [0.14]
	(-1.55)[0.12]	(-0.77)[0.44]				
3	0.81	-2.07	-3.18			0.92 [0.43]
	(0.28)[0.78]	(-0.71)[0.48]	(-1.2)[0.23]			
4	-0.94	4.17				2.9 [0.06]
	(-0.46)[0.64]	(2.38)[0.02]				
5	-0.78	-1.1	-2.57			1.59 [0.19]
	(-0.49)[0.62]	(-0.69)[0.49]	(-1.64)[0.1]			
6	-0.609	0.268	-1.07			0.43 [0.73]
	(-0.55)[0.58]	(0.23)[0.81]	(-1.01)[0.31]			
7	0.405	-0.143	-0.026	0.644	-0.702	0.164 [0.98]
	(0.398)[0.69]	(-0.13)[0.9]	(-0.024)[0.98]	(0.604)[0.55]	(-0.575)[0.56]	

z-statistic in parentheses, p-value in brackets.
F-Test is a Wald-Test with the null hypothesis that all coefficients are jointly zero.
Dependent Variable *Democracy Score in t Conditional on Democracy Score in* $(t-1)$.
$N = 128, 1973 - 2003$, Freedomhouse Data.

Table 2.6. Ordered Logit with 7 Categories.

Democracy Score in $(t-1)$	Growth in $(t-1)$	Growth in $(t-2)$	Growth in $(t-3)$	F-Test
1	-9.3	1.3	-7.4	4.5 [0.004]
	(-2.7)[0.008]	(0.3)[0.74]	(-2.1)[0.04]	
2	-1.08	1.003	-3.61	2.81 [0.04]
	(-0.7)[0.49]	(0.68)[0.49]	(-2.7)[0.007]	
3	-0.35	1.69	-1.6	1.54 [0.2]
	(-0.34)[0.73]	(1.76)[0.08]	(-1.54)[0.12]	

z-statistic in parentheses, p-value in brackets.
F-Test is a Wald-Test with the null hypothesis that all coefficients are jointly zero. Dependent Variable *Democracy Score in t Conditional on Democracy Score in* $(t-1)$. $N = 128, 1973-2003$, Freedomhouse Data. Categories: $\{1,2\} = 1, \{3,4,5\} = 2, \{6,7\} = 3$.

Table 2.7. Ordered Logit with 3 Categories.

Democracy Score in $(t-1)$	Growth in $(t-1)$	Growth in $(t-2)$	Growth in $(t-3)$	F-Test
1	-7.54	7.46	-6.61	2.48 [0.06]
	(-1.94)[0.05]	(1.7)[0.09]	(-1.74)[0.08]	
2	-0.36	-0.88	-1.34	0.62 [0.6]
	(-0.23)[0.81]	(-0.58)[0.56]	(-0.93)[0.35]	
3	-1.06	0.74	-1.18	0.82 [0.48]
	(-1.007)[0.31]	(0.72)[0.47]	(-1.11)[0.27]	

z-statistic in parentheses, p-value in brackets.
F-Test is a Wald-Test with the null hypothesis that all coefficients are jointly zero. Dependent Variable *Democracy Score in t Conditional on Democracy Score in* $(t-1)$. $N = 128, 1973-2003$, Freedomhouse Data. Categories: $1 = \textit{Free}, 2 = \textit{Partly Free}, 3 = \textit{Not Free}$

Table 2.8. Ordered Logit with 3 Categories.

Democracy Score in $(t-1)$	Growth in $(t-1)$	Growth in $(t-2)$	Growth in $(t-3)$	F-Test
1	-22.46	-6.7	-20.98	3.94 [0.008]
	(-2.13)[0.03]	(-0.59)[0.56]	(-1.99)[0.05]	
2	-4.45	2.44	9.07	1.3 [0.27]
	(-0.78)[0.43]	(0.43)[0.66]	(1.73)[0.08]	
3	0.86	6.07	11.22	2.6 [0.05]
	(0.18)[0.86]	(1.42)[0.15]	(2.47)[0.01]	

z-statistic in parentheses, p-value in brackets.
F-Test is a Wald-Test with the null hypothesis that all coefficients are jointly zero. Dependent Variable *Democracy Score in t Conditional on Democracy Score in* $(t-1)$. $N = 43, 1950 - 2006$, Polity IV Data.
Categories: $\{7, ..., 10\} = 1, \{-6, ..., 6\} = 2, \{-10, ..., -7\} = 3$

Table 2.9. Ordered Logit with 3 Categories.

Independent Variable	$\triangle y_{it}$	$\triangle y_{it}$	$\triangle y_{it}$
Constant	0.002	0.003	0.004
	(1.83)[0.07]	(2.38)[0.02]	(2.98)[0.003]
$\triangle y_{it-1}$	0.21	0.2	0.23
	(11.8)[0]	(11.96)[0]	(14.3)[0]
$\triangle y_{it-2}$	0.02	0.02	
	(1.24)[0.21]	(1.3)[0.2]	
$\triangle y_{it-3}$	0.07	0.07	
	(3.86)[0]	(4.58)[0]	
$\triangle y_{it-4}$	0.01		
	(0.67)[0.5]		
$\triangle y_{it-5}$	-0.004		
	(-0.26)[0.8]		
Category 1	0.003	0.003	0.003
	(1.97)[0.05]	(1.94)[0.05]	(1.79)[0.39]
Category 2	0.002	0.002	0.002
	(1.2)[0.23]	(1.06)[0.28]	(0.84)[0.39]
Category 3	0.001	0.001	0.001
	(0.6)[0.55]	(0.53)[0.6]	(0.65)[0.52]
Category 4	0.003	0.003	0.003
	(1.2)[0.23]	(1.33)[0.18]	(1.16)[0.25]
Category 5	-0.001	-0.0009	-0.001
	(-0.64)[0.52]	(-0.46)[0.64]	(-0.7)[0.49]
Category 6	-0.002	-0.002	-0.002
	(-1.1)[0.29]	(-1.21)[0.22]	(-1.03)[0.3]
Schwarz criterion	-4.14	-4.15	-4.11
F-Test	2.51 [0.02]	2.52 [0.02]	2.13 [0.05]

Dependent Variable $\triangle y_{it}$.
t-statistic in parentheses, p-value in brackets.
F-Test is a Wald-Test with the null hypothesis that all category dummys are jointly zero.
$N = 128, 1977 - 2003$, Freedomhouse Data.

Table 2.10. Estimates for Panel OLS, $K = 7$.

Independent Variable	$\triangle y_{it}$	$\triangle y_{it}$	$\triangle y_{it}$
Constant	0.001	0.002	0.003
	(1.56)[0.12]	(2.17)[0.03]	(3.26)[0.001]
$\triangle y_{it-1}$	0.21	0.2	0.23
	(11.83)[0]	(11.98)[0]	(14.3)[0]
$\triangle y_{it-2}$	0.02	0.02	
	(1.24)[0.21]	(1.3)[0.2]	
$\triangle y_{it-3}$	0.07	0.07	
	(3.84)[0]	(4.51)[0]	
$\triangle y_{it-4}$	0.01		
	(0.61)[0.54]		
$\triangle y_{it-5}$	-0.006		
	(-0.35)[0.72]		
Category 1	0.004	0.004	0.003
	(3.21)[0.01]	(3.21)[0.001]	(2.79)[0.01]
Category 2	0.002	0.002	0.001
	(1.23)[0.21]	(1.47)[0.14]	(1.13)[0.26]
Schwarz criterion	-4.148	-4.16	-4.12
F-Test	5.22 [0.01]	5.15 [0.01]	3.92 [0.02]

Dependent Variable $\triangle y_{it}$.
t-statistic in parentheses, p-value in brackets.
F-Test is a Wald-Test with the null hypothesis that all category dummys are jointly zero.
$N = 128, 1977 - 2003$, Freedomhouse Data.
Categories: $\{1, 2\} = 1, \{3, 4, 5\} = 2, \{6, 7\} = 3$

Table 2.11. Estimates for Panel OLS, $K = 3$.

Independent Variable	$\triangle y_{it}$	$\triangle y_{it}$	$\triangle y_{it}$
Constant	0.002	0.002	0.003
	(1.93)[0.05]	(2.48)[0.01]	(3.39)[0]
$\triangle y_{it-1}$	0.21	0.2	0.23
	(11.84)[0]	(11.97)[0]	(14.32)[0]
$\triangle y_{it-2}$	0.02	0.02	
	(1.26)[0.21]	(1.3)[0.19]	
$\triangle y_{it-3}$	0.07	0.07	
	(3.85)[0]	(4.51)[0]	
$\triangle y_{it-4}$	0.01		
	(0.63)[0.53]		
$\triangle y_{it-5}$	-0.006		
	(-0.32)[0.74]		
Category 1	0.004	0.004	0.003
	(2.84)[0.005]	(2.89)[0.004]	(2.63)[0.009]
Category 2	0.0005	0.0007	0.0005
	(0.38)[0.71]	(0.58)[0.56]	(0.39)[0.69]
Schwarz criterion	-4.148	-4.16	-4.12
F-Test	4.94 [0.01]	4.76 [0.01]	4.09 [0.02]

Dependent Variable $\triangle y_{it}$.
t-statistic in parentheses, p-value in brackets.
F-Test is a Wald-Test with the null hypothesis that all category dummys are jointly zero.
$N = 128, 1977 - 2003$, Freedomhouse Data.
Categories: $1 = \textit{Free}, 2 = \textit{Partly Free}, 3 = \textit{Not Free}$

Table 2.12. Estimates for Panel OLS, $K = 3$.

Independent Variable	$\triangle y_{it}$	$\triangle y_{it}$	$\triangle y_{it}$
Constant	0.006	0.007	0.008
	(5.41)[0]	(6.07)[0]	(6.97)[0]
$\triangle y_{it-1}$	0.2	0.17	0.18
	(9.49)[0]	(8.16)[0]	(8.94)[0]
$\triangle y_{it-2}$	-0.02	-0.01	
	(-1.21)[0.23]	(-0.72)[0.47]	
$\triangle y_{it-3}$	0.1	0.1	
	(4.94)[0]	(4.87)[0]	
$\triangle y_{it-4}$	-0.01		
	(-0.51)[0.61]		
$\triangle y_{it-5}$	0.04		
	(2.1)[0.04]		
Category 1	0.0003	0.0004	0.0003
	(0.24)[0.81]	(0.34)[0.74]	(0.26)[0.8]
Category 2	0	-0.0004	-0.0008
	(-0.04)[0.96]	(-0.29)[0.77]	(-0.64)[0.52]
Schwarz criterion	-5.26	-5.24	-5.21
F-Test	0.1 [0.91]	0.43 [0.65]	0.91 [0.4]

Dependent Variable $\triangle y_{it}$.
t-statistic in parentheses, p-value in brackets.
F-Test is a Wald-Test with the null hypothesis that all category dummys are jointly zero.
$N = 43, 1950 - 2006$, Polity IV Data.
Categories: $\{7, ..., 10\} = 1, \{-6, ..., 6\} = 2, \{-10, ..., -7\} = 3$

Table 2.13. Estimates for Panel OLS, $K = 3$.

ADF t-statistics, 1973 - 2003

Afghanistan	-0.4332	Germany	-0.8995	Norway	-0.3524
Algeria	-2.0264	Ghana	-0.7829	Oman	-1.9844
Argentina	-2.1335	Greece	0.8433	Pakistan	-1.5355
Australia	1.7648	Guatemala	-2.4451	Panama	-1.2485
Austria	0.1972	Guinea	0.0289	Paraguay**	-3.1092
Bahrain	-1.9983	Honduras**	-3.9733	Peru	-2.3541
Bangladesh	2.7225	Hungary	0.0413	Philippines	-2.0161
Barbados	-1.2454	Iceland	-0.5333	Poland	-0.9100
Belgium	0.6742	India	2.7605	Portugal	-0.2900
Benin	-1.1529	Indonesia	-0.2001	Qatar**	-5.9831
Bhutan	0.5184	Iran	-1.8309	Romania**	-3.3787
Bolivia	-1.7267	Iraq	-2.2460	Rwanda	-1.5569
Botswana	0.3283	Ireland	0.9199	Samoa	-0.9261
Brazil	-2.4545	Israel	-0.2712	Saudi Arabia	-1.9816
Brunei	-1.3849	Italy	-0.5056	Senegal**	-3.5265
Burkina Faso	0.4748	Jamaica	-1.4809	Sierra Leone	-0.2130
Burundi	-1.5015	Japan	-1.6479	Singapore	-0.6851
Cambodia	-2.1378	Jordan	-2.0096	Somalia	-1.5992
Cameroon	-2.0165	Kenya	-2.1580	South Africa	-0.8579
Canada	0.1562	Korea, Dem. Rep.	-2.2434	Spain	1.0928
Central African Republic	-1.5943	Korea, Republic of	1.2363	Sri Lanka	1.4201
Chad	-2.4618	Kuwait*	-2.6424	Sudan	-2.5250
Chile	0.3678	Laos	-0.5890	Swaziland	-0.7851
China	4.1082	Lesotho	0.1335	Sweden	0.1869
Colombia	-1.2897	Liberia	-1.0431	Switzerland	-0.6647
Congo, Dem. Rep.	-1.6907	Luxembourg	0.8728	Syria	-2.2540
Congo, Republic of	-1.6453	Madagascar	-1.7206	Taiwan	0.9980
Costa Rica	-0.2350	Malawi	-2.1550	Tanzania	-0.0385
Cote d'Ivoire**	-3.0152	Malaysia	1.2191	Thailand	-0.2899
Cuba	-1.9333	Maldives	1.3144	Togo	-1.3878
Cyprus	0.0507	Mali	0.0411	Tonga	-1.6026
Denmark	0.2111	Malta	-1.2806	Trinidad & Tobago	0.2704
Dominican Republic	0.1744	Mauritania	-1.2154	Tunisia	3.1829
Ecuador	-2.5457	Mauritius	1.3489	Turkey	-0.7587
Egypt	0.4926	Mexico	-1.8017	Uganda	-0.8921
El Salvador	-1.8230	Mongolia*	-2.7874	United Arab Emirates	-1.8188
Equatorial Guinea	0.0251	Morocco	-1.9919	United Kingdom	0.5254
Ethiopia	-0.0744	Nepal	0.3829	United States	0.0407
Fiji	-2.4159	Netherlands	-0.2195	Uruguay	-1.9097
Finland	-0.5145	New Zealand	1.6248	Venezuela	-0.8928
France	0.1434	Nicaragua	-1.2328	Zambia	-0.7498
Gabon	-1.8599	Niger	-1.3281	Zimbabwe	-2.5500
Gambia, The	-2.5285	Nigeria	-2.1138		

5% and 10% critical values, −2.9370 and −2.6152
* Significant at the 5% level.
** Significant at the 10% level.
Null: A unit root is present.

Table 2.14. ADF t-statistic (one lag, intercept) for all countries.

Chapter 3

The Effect of Extending the Franchise to Women: The Case of Pittsburgh

3.1 Introduction

There is an extensive literature on the extension of the franchise. The motivation behind this literature is best formulated by Przeworski, Alvarez, Cheibub, and Limongi (2000).

> Why would people who monopolize political power ever decide to put their interests or values at risk by sharing it with others? Specifically, why would those who hold political rights in the form of suffrage decide to extend these rights to anyone else?

The most popular explanation stresses the power of the disenfranchised threatening those holding political power with social unrest.[1] This explanation is motivated by the extension of the franchise to poor workers in the Western world in the nineteenth century. While there were clearly instances of social unrest credibly supporting the threat by the poor workers there are no instances known where women would organize a violent uprising in order to push through a redistribution of political power. This makes the "revolution hypothesis" an inferior candidate for explaining the extension of the franchise to women.

Poor workers and the rich elite would fight over the distribution of power because the latter would determine the distribution of income. In the case of women suffrage, it is not even clear if or why men and women would argue over the distribution of power. Suppose for example, husbands and wives have identical interests across social groups, i.e., male and female workers as well as male and female members of the elite have similar preferences, respectively. In that case extending the franchise to women would have no effect on election results since the number of votes for each party or candidate would simply double. If that were true empirically, it would raise the question why women had to fight for their suffrage. If however, women vote differently than men than this might explain the fight for women suffrage. In fact, different turnout rates across women might have a significant effect. Suppose there are two significantly different groups in the country, say A and B where men and women within each group have similar preferences. Furthermore, suppose party PA represents A's interests while party PB represents B's interests. If, after the extension of the franchise to women, the women of group A have a higher turnout rate at the polls than the women of group B than this would explain PA supporting women suffrage while PB is opposing it.

[1] See for example Acemoglu and Robinson (2000).

This paper investigates empirically the effect of extending the franchise to women. I focus on two questions. First: Is female voter turnout different across different social groups? In particular, I investigate whether richer and more educated women have a higher turnout rate at the polls than poorer and less educated women. The second question is: Does extending the franchise to women alter the election results significantly? The motivation is simple. Even if a higher percentage of richer and more educated women is voting than poorer and less educated women, it is possible that the election results shift in favor of the poorer districts if the total number of poorer women is greater the than the total number of richer women voting. Also, it is possible that women vote homogenously as a group since they might have different preferences than men. In the former case, wives vote as their husbands while in the latter case women vote differently than men.

The question in the background is why women had to fight for their right to vote. In order to understand this question better it seems appropriate to look at the immediate effect of extending the franchise to women. This approach contrasts with work interested in female voting preferences looking at the so-called gander gap in recent elections. In particular, I focus on Pittsburgh. There are two reason for doing so. First, Pennsylvania was forced in 1920 to grant women the right to vote. This means women suffrage was less of an endogenous phenomenon and more of an exogenous shock. Second, I have access to data on a ward level for Pittsburgh. Pittsburgh has 27 wards. I have demographic and voting data before and after women suffrage was enacted for each ward. Also, Pittsburgh has enough social variation to investigate whether women suffrage has different effects across different social groups. Additionally, I have voter registration data by gender. This leaves me with a solid empirical basis to investigate the questions of interest.

I have three sets of results. First, I interpolate the number of male voters in the presidential election of 1920 from the presidential election of 1916 and back out the number of female voters per ward. Qualitative evidence on the location of poor and rich neighborhoods in Pittsburgh at that time suggests that female voter turnout and female voter participation were higher in richer and more educated neighborhoods. Second, I regress female voter turnout and female voter participation on a number of variables indicating social status. I find that the number of foreign-born men has a negative effect on female voter turnout and female voter participation. Male immigrants often came first alone before bringing their wives over. Since these immigrants were often poor and had left school early I conclude that female voter turnout and female voter participation was higher in the richer parts of town. The next question is whether the enactment of women suffrage had a significant effect on the election results. To this end, I regress the share of Republican votes in 1920 on female voter turnout and female voter participation. I cannot find a significant effect of these variables on the share of Republican votes. There might be a number of reasons why there is no effect. There might have been election-specific issues which dominated the effect of the enactment of women suffrage. On the other hand, the influence of the local Republican party in Pittsburgh might have been too dominant to allow for a significant effect of women suffrage. Third, I look at voter registration data from 1925 and the early 1930s. Qualitative evidence on the location of poor and rich neighborhoods suggests that more female voters registered in rich neighborhoods than in poor neighborhoods in 1925. Also, comparing the female voter participation rates from 1920, 1925 and the early 1930s suggest that female voter participation was rising after the initial enactment of women suffrage confirming the hypothesis that it takes some time for a newly enfranchised group to show up at the polls. I view this work as a pilot study. The reason is that I only discuss the voter registration data qualitatively. Future work would record the voter registration data much more carefully, so that these data could be used for statistical and econometric inference.

The outline of this paper is as follows. First, I review the literature. In section 3.3 and 3.4 I discuss briefly the history of women suffrage in the U.S. and the history of Pittsburgh, respectively. Section 3.5 gives a description of the data while the analysis is conducted in sections 3.6 and 3.7.

3.2 Literature Review

The main problem with empirically investigating the effect of extending the franchise to women is the availability of data. When the franchise was extended to women in America in the late nineteenth and early twentieth century there were no records kept on voter registration or vote by gender. Similarly, there where no exit polls back then. There are a number of case studies. Berman (1993) investigates the elections in Arizona in 1914 and 1916. The percentage of female registered voters in a given election precinct is approximated by the percentage of women in the total population. The results suggest that women relative to men appear to oppose progressive labor reform, the abolition of the death penalty and favor prohibition. In general women appear to be more conservative than men as they tend to support the Republican and Progressive party while opposing the Socialist party. Ogburn and Goltra (1919) find similar results for Oregon in 1914. Their results suggest that women relative to men oppose the eight-hour day for women, proportional representation, the abolition of the senate and spending public money. Goldstein (1984) on the other hand, looks at Chicago during the period between 1914 and 1921. The great advantage of his study is that votes were counted separately for men and women. He cannot find a significant effect, suggesting that both genders vote similarly.

Then, there is a related empirical literature trying to link the introduction of women suffrage and the direction of government policy. The idea behind looking at government policy instead of election results is that it might take some time for the full voter turnout of the newly enfranchised group. Lott and Kenny (1999) estimate the female voter turnout and empirically connect this additional turnout to increased government expenditure and revenue as well as the more liberal voting record in Congress. The theoretical explanation for this phenomenon is that married women do not accumulate enough labor market-specific human capital which makes them more reliable on government support after a divorce. Miller (2008) shows empirically that the extension of the franchise to women in America lowers child mortality by $8-15\%$. The idea is that women care more about children than men. The empirical identification strategy is to use variation across states which have granted women the right to vote at various times. Funk and Gathmann (2006) investigate the introduction of women suffrage in Switzerland in 1971. They use the fact that Switzerland has many opportunities for democratic participation. In particular, they use detailed survey data on a large number of federal referendums. They find that the enactment of women suffrage changed the scope of government away from spending on defense and agricultural subsidies towards spending on public transportation and environmental causes. There is also an empirical literature in political science on women suffrage and female voter turnout. Corder and Wolbrecht (2006) for example, use a Bayesian approach to ecological inference for the presidential elections of 1920 and 1924. They find that initially female voter turnout is lower than male voter turnout. Over time however, the same factors that determine male voter turnout also determine female voter turnout.

There are not many theoretical explanations as to why men would share power with women voluntarily. Tertilt and Doepke (2009) describe the following tradeoff. Sharing their power lowers men's intra-household bargaining power. On the other side, sharing their power increases men's daughters' intra-household bargaining power with regard to their husbands. Also, an increase in women's rights translates into an increase in human capital of the men's children. Then, an increase in the returns to education explains the increase in women's rights. Next, I give a brief review of the history of women suffrage.

3.3 The History of Women Suffrage in the U.S. and Pennsylvania

The official beginning of the history of women suffrage in the United States is the Seneca Falls Convention 1848.[2] It produced the Declaration of Sentiments demanding among other things the right to vote for women. The issue of women suffrage gained publicity again after the end of the civil war. The proposed fifteenth amendment to the constitution gave African-American men the right to vote. The women suffrage movement split over the question whether to support or oppose the fifteenth amendment. Both factions reunited in 1890 to form the National American Women Suffrage Organization providing an umbrella organizations for the numerous clubs and local organizations throughout the country. States in the West started to give women the right to vote in the late nineteenth century. Colorado is the first state granting the right to vote to women in 1893, Utah and

[2]Flexner and Fitzpatrick (1996) is a standard reference on the subject.

Idaho follow in 1896 and Washington in 1910. Figure 3.1 on page 67 gives on overview of the development across states.

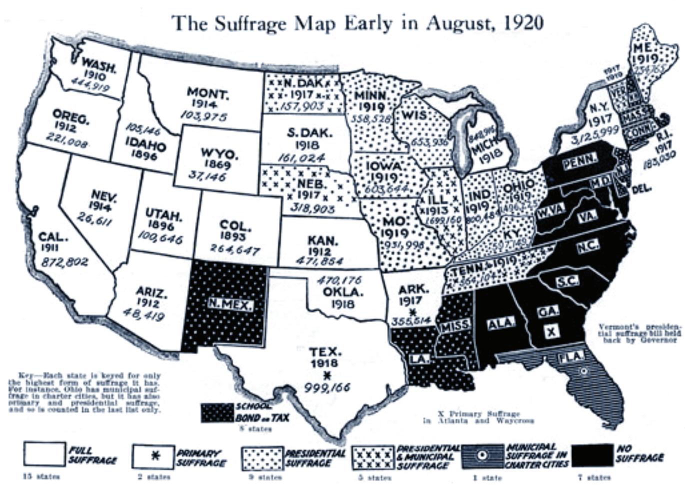


Figure 3.1. The Development of Woman Suffrage in the U.S.

World War I pushed other issues on top of the political agenda. However, within the official line of reasoning that the United States would make the world safe for democracy the women suffrage movement demanded to make America a full democracy at home as well. Congress passed the nineteenth amendment to the constitution in the summer 1920 making the presidential elections in the fall of 1920 the first elections were all women older than 21 in the United States were allowed to vote.

The congressmen from Pennsylvania in the House of Representatives voted mainly against the nineteenth amendment. Even though Pennsylvania was one of the first states where the women suffrage movement was organized it was one of the last states that was forced by Congress to grant women the right to vote. Even in 1915 the women suffrage movement lost a referendum to extend the franchise to women in Pennsylvania. In fact, by 1913 nine states had granted women full suffrage and twenty-one states had granted partial suffrage. In Pennsylvania however, women were fully disenfranchised and not even allowed to vote on school matters.[3] The next section discusses the history and social geography of Pittsburgh.

3.4 The History and Social Geography of Pittsburgh

The main task of this work is to understand whether different social groups had a different attitude towards extending the franchise to women. To this end, I use geographical data for the city of Pittsburgh arguing that certain social groups live in certain parts of town. This section provides the background on the geographical location of various groups.

Pittsburgh was a fast growing city in the nineteenth century. In 1800 it had 1,565 inhabitants, in 1850 46,601 and in 1900 321,616 inhabitants. There are a number of reasons for this growth. First, Pittsburgh is located at the intersection of three rivers, the Ohio, the Allegheny and the Monongahela river. This made it a natural hub for the growing commerce. Second, the region surrounding Pittsburgh had an abundance of natural resources, such as timber, ore, clay, steel, coal and natural gas.[4] This put Pittsburgh in an ideal position to participate in the industrial rise of the nineteenth century. In the middle of the nineteenth century Pittsburgh was a major producer of steel, iron, glass, textile products, coke and railroad equipment. Pittsburgh was transformed massively starting in the 1870s due to technological progress in its major industries. The introduction of the Bessemer process and the development of more productive blast furnaces in the steel

[3]See Roessing (1914).

[4]See Muller (2001) for a detailed description.

production for example, allowed for the creation of large vertically integrated steel mills. Steamboats and the railroad would allow the creation of large industrial sites on flood plains within and outside of Pittsburgh. This technological progress led to a decreasing demand for craftsmen skilled in the production of steel and a rising demand for unskilled labor. This demand was mainly satisfied by an increasing number of immigrants who were willing to work for falling wages in the steel mills.[5] These immigrants lived typically close to the production sites. Another reason for large production sites moving out of town was to weaken the influence of the unions. The most famous labor dispute of the time was the Homestead strike in 1892 which resulted in a victory for the Carnegie Steel company. Muller (2001) concludes:

> By 1910 the Pittsburgh district was a complex urban landscape with a dominant central city, surrounded by proximate residential communities, mill towns, satellite cites, and hundred of mining towns.

Koppes and Norris (1985) provide an interesting study highlighting the social geography of Pittsburgh at the turn of the century. They study typhoid fever in Pittsburgh between 1890 and 1910. The hypothesis underlying their work is that typhoid fever is causing high mortality rates mainly among the working class while sparing the upper class. Typhoid fever is contracted mainly via polluted water. This means digging wells, buying water and filtering or boiling water reduces the exposure to typhoid fever significantly. Obviously, these protective measures favor those who have the money or time to comply with them. Pittsburgh had the highest typhoid mortality rate in the nation at the turn of the century. The reason was that the city was still relying on its three rivers to provide the population with clean

[5]See Bodnar, Roger, and Weber (1982) for a more detailed description.

water when those rivers were actually part of the problem. Only in 1908 the city started to provide a public filtration system leading to a sharp decline in typhoid fever mortality rates. Koppes and Norris (1985) look at the part of town between the Allegheny and the Monongahela river and the part of town south of the Monongahela river. Both parts of town lead to similar conclusions: The neighborhoods close to the river, i.e. the flood plains, had the highest mortality rates whereas in the upland the mortality was fairly moderate.[6] Since the large steel mills and other production sites were mostly located near the rivers the unskilled workers were forced to live in the flood plain parts of town. This leads Koppes and Norris (1985) to conclude that working class districts had a significantly higher mortality rate than upper class districts. Interestingly, the authors observe that certain hill districts which showed low mortality rates in the 1890s would show high mortality rates in 1907. The authors note that these hill districts were mostly close to the flood plain districts suggesting that the middle class moved farther away from the river leading to new immigrants overtaking those districts. In other words, the disease would follow the poor. Their general conclusion that mortality rates were significantly higher in working class districts in the flood plains is complemented by the qualitative study by Bodnar, Roger, and Weber (1982) who report that the flood plain districts were mainly populated by poor immigrants while the upper class lived up the hills.

From a political perspective, Pittsburgh was a Republican stronghold at the end of the nineteenth and during the early twentieth century. Table 3.1 on page

[6]First, they look at the area between the Allegheny and the Monongahela river. Of the 9 wards in the top quartileof mortality due to typhoid fever, 6 were in the flood plain districts. In the part of town south of the Monongahela river the numbers are as follows. The average mortality rate per 100,000 inhabitants between 1892 and 1899 is in the flood plain districts 86.3, 51.0, 37.1, 38.3, 68.1, 75.6, 129.4 and 50.8. The numbers in the upland districts are 60.5, 37.7, 65.2, 39.4 and 61.2.

Year	Candidate (D)	Candidate (R)	Votes (D)	Votes (R)
1896	Bryan	McKinley*	12,925	32,409
1900	Bryan	McKinley*	12,184	29,647
1904	Parker	Roosevelt*	10,871	38,728
1908	Bryan	Taft*	15,847	29,925
1912	Wilson*	Roosevelt	17,352	25,394
1916	Wilson*	Hughes	26,947	42,059
1920	Cox	Harding*	20,456	74,275
1924	Davis	Coolidge*	9,814	64,795
1928	Smith	Hoover*	84,129	91,041
1932	Roosevelt*	Hoover	88,886	64,134
1936	Roosevelt*	Landon	173,538	72,223
1940	Roosevelt*	Willkie	187,288	116,980
1944	Roosevelt*	Dewey	159,938	103,001
1948	Truman*	Dewey	160,780	102,301

An asterisk shows the winner.
Votes (R) in 1912 are the sum of Bull Moose, Progressive and Washington. The Republican candidate received 14,658 votes.

Table 3.1. Presidential Election Results in Pittsburgh.

71 shows the total number of votes for the Republican and Democratic presidential candidate, respectively. The Republican candidate received an overwhelming majority of the vote until the mid-1920s. Only Roosevelt and the challenging economic times were able to turn the tide in the early 1930s. One of the reasons for the strong voting in favor of the Republican candidates was a culture of patronage installed by the local Republican party.[7] The so-called machine was particularly effective with respect to immigrant groups. According to Luconi (1996), in 1926 the Republican candidate for governor, John S. Fischer, received 91.6% of the Italian-American votes in Pittsburgh.

Interestingly, the machine prompted reformers to enact strict voting laws in

[7]Between 1880 and 1930 only two mayors were Democrats.

Pennsylvania by 1906. Reformers were highly concerned that the machine was receiving too many votes due to voting fraud. Also, the well-educated reformers from the East end of Pittsburgh were suspicious of the working class voting since they suspected that those would vote for whomever they had been told by their corporate superior without having the public good at heart. Two quotes from the historic novel *Out of this Furnace* by Bell (1950) are presented in the appendix on page 88 showing the influence of the company over how workers voted.[8] Mayfield (1993) summarizes the personal registration law of 1906 as follows:

> Eligible males had to go to their district polling place on one of three specific days at least five weeks before the election and prove that they were over twenty-one years old, a native-born or naturalized foreign-born citizen, and had paid a state or county tax within the previous two years. As under the nonpersonal system, adult males without property were supposed to pay an occupation tax to be allowed to vote, ranging from twenty-five to fifty cents.

The well-educated reformers were pleased with these new restrictions as they suspected that this would lead to only sincere voters showing up at the polls.

3.5 Data Description

I have access to the lists of registered voters per ward and election district for various years provided by the Pittsburgh Board of Registration Committee Records (1906–1936). These data are discussed in detail in section 3.7. The voting records for the wards of Pittsburgh for the presidential elections are from Smull (1897–1949). All other data are from the Census (1910) and Census (1920). For the

[8]This novel is also used as a reference by Barrett and Roediger (1997).

empirical analysis it is most important to have some variation across wards with respect to social composition. This is particularly important as I do not have demographic data on a precinct level of which there are approximately 300 in Pittsburgh. Instead I have only 27 wards. Table 3.2 on page 74 provides summary statistics for a number of relevant variables. First, Pittsburgh's population was still growing between 1910 and 1920. Second, there is a significant variation of the ratio of men to women across wards. A comparison between wards 2 and 14 illustrates the contrasts that are substantiated more generally by the statistical analysis that follows. For example, in 1920 there was a male-to-female ratio of 1.32 in ward 2. According to figure 3.5 on page 100, ward 2 is located on the river where the iron mills were producing suggesting that this is a poor neighborhood where a lot of male immigrants lived which came initially without their families. On the other hand, ward 14 had a male-to-female-ratio of 0.78. Ward 14 is composed of richer neighborhoods in the east of Pittsburgh since the households in that area could afford to have female staff living with them. Similarly, the illiteracy rate was 19% in ward 2 while it was 2% in ward 14. This pattern is confirmed for the ratio of foreign born men relative to the male population. In ward 2, 40% of all men were born abroad while in ward 14 13% of all men were born abroad. Homes owned per ward is a good indicator for the level of wealth. 4% of all homes were owned in ward 2 while 43% of all homes were owned in ward 14. Also, voter turnout itself is an indicator for wealth because Pennsylvania had a poll tax until 1933. This makes it less surprising that the voter turnout in ward 2 was 25% while it was 41% in ward 14. The standard deviations for all of these variables in table 3.2 are quite large suggesting the is a good amount of variation across wards to be exploited.

	1910	1920
Av. Population per Ward	19,700	21,567
	[4,248]	[4,830]
Av. Ratio of Men to Women per Ward	1.05	1.02
	[0.17]	[0.12]
Av. Illiteracy Rate per Ward	5.9	4.6
	[5.3]	[3.7]
Av. % of Foreign-Born Men per Ward		22.2
		[7.9]
Av. Voter Turnout per Ward	41.2	28.7
	[9.9]	[6.1]
Av. % of Homes Owned per Ward		26.9
		[11.6]

Standard deviation in brackets.
Pittsburgh has 27 Wards.
Illiteracy rate is the percentage of illiterates in the population older than 10 years.
Foreign-born men relative to all men.
Voter turnout is the total number of votes relative to the number of eligible voters.
Voter turnout in 1910 means 1912.
In 1912 men over 21 are eligible voters.
In 1920 women and men over 21 are eligible voters.

Table 3.2. Summary Statistics for Pittsburgh.

3.6 Statistical Model and Empirical Results

There are 27 wards denoted by i, $i = 1, ..., 27$ and three time periods $t_0 = 1910, t_1 = 1916, t_2 = 1920$. The male (female) population older than 21 in year t in ward i is denoted by $M_i(t)$ ($F_i(t)$). I observe $M_i(t_0)$ and $M_i(t_2)$ but not $M_i(t_1)$ because 1916 is not a Census year. I approximate the number of eligible voters in 1916 $\widehat{M_i}(t_1)$, i.e., the male population older than 21, via linear interpolation:

$$\widehat{M_i}(t_1) = M_i(t_0) + \frac{M_i(t_2) - M_i(t_0)}{10} \times 6.$$

I use this approximation to estimate the voter turnout in the presidential elections of 1916 which was the last presidential election in which women in Pennsylvania were refused to vote. Define the male voter turnout in ward i in year t, $T_i^m(t)$ as the percentage of men voting relative to all males older than 21 and denote the overall number of votes per ward i in year t by $V_i(t)$. Then, male voter turnout in ward i in 1916 is given as follows:

$$\widehat{T}_i^m(t_1) = \frac{V_i(t_1)}{\widehat{M}_i(t_1)} \times 100.$$

In order to estimate the female voter turnout in 1920 I make the assumption that male voter turnout is the same in 1916 and 1920.

Assumption 4 *For all wards i,* $T_i^m(t_1) = T_i^m(t_2)$.

Using this assumption I estimate female voter turnout per ward i in 1920, $T_i^f(t_2)$ as

$$\widehat{T}_i^f(t_2) = \left(V_i(t_2) - \frac{\widehat{T}_i^m(t_1) M_i(t_2)}{100} \right) \times \frac{100}{F_i(t_2)}.$$

Also, in order to understand whether there is variation in female voter turnout across wards I calculate the percentage of female voters relative to all voters per ward in 1920,

$$\frac{\widehat{T}_i^f(t_2) F_i(t_2)}{V_i(t_2)}.$$

I call this statistic female voter participation rate. The empirical results can be found in table 3.3 on page 90. Comparing those results with the location of the wards as shown in figure 3.5 on page 100 suggests that female voter turnout was higher in richer neighborhoods. For example, female voter turnout was particularly low in wards 1,3,16,17,23 and 24 which are close to the rivers where the steel mills were located. Wards 7,8,11 and 14 which are located in the East on the hills had a particularly high female voter turnout. Also, the female voter participation rate varies significantly across wards. If women in every ward would have had identical turnout rates to men in the respective ward, then we should see that 50% of all voters were women. However, in wards 16 and 24 only 5% of all voters were women. Both wards are close to the rivers suggesting they are poor neighborhoods. In contrast, the rich wards up the hill 7 and 8 have the highest percentage of female voters with 43% and 40%, respectively.

Next, I regress female voter turnout and female voter participation rate on a number of measures for wealth and education such as the illiteracy rate per ward, the ratio of men to women per ward, foreign-born men per ward and the percentage of homes owned per ward. Except for the ratio of men to women per ward, all variables are percentages bound between 0 and 100. In order to specify a tractable statistical model by assuming an unbounded error distribution, I transform these variables monotonically, so that they are stretched out over the real line. That is, instead of working with x, I work with $\ln\left(\frac{x}{100-x}\right)$.

First, I look at the scatter diagrams for female voter turnout on page 91. The illiteracy rate and female voter turnout have a negative relationship. This is in line with the main hypothesis that female voter turnout is higher in richer wards if one assume that illiteracy is higher in poorer wards. Note however, that there is an outlier which could have a significant effect given that there are only 27 wards. While most wards have an illiteracy rate between 2% and 8%, ward 2 has an illiteracy rate of 19%. Clearly, ward 2 is a working class ward near the river. The adjacent ward 6 has an illiteracy rate of 9%. In 1910, wards 2 and 6 had illiteracy rates of 17% and 24%, respectively. In the analysis I control for the outlier. The ratio of men to women and the ratio of foreign-born men have a negative relationship with female voter turnout in line with the main hypothesis. The more men relative to women a ward had, the more likely it is that it is a working class ward. The are two reasons for this. First, male immigrants would come first followed by their wives. Second, richer households up the hills had female staff members living with them. Similarly, a large number of foreign born men indicates a large number of immigrants which would generate a poor neighborhood. Then, the higher the number of homeowners the higher the female voter turnout suggesting that female voter turnout is higher in richer wards. Since there was a poll tax in Pennsylvania at the time, I hypothesize that voter turnout in 1912 might have been higher in richer neighborhoods. The picture suggests that voter turnout in 1912 and female voter turnout have a positive relationship suggesting that female voter turnout is higher is richer neighborhoods.

Model 1 in table 3.4 on page 92 includes all these five regressors. Note first that illiteracy rate and homes owned have the wrong sign compared to the scatter diagrams on page 91 which agree with the main hypothesis. One possible explanation is that the outlier ward 2 produces this result. In order to control for ward 2 I add a dummy variable for ward 2. The results are given in the column for model 2. The sign for homes owned and illiteracy rate remain wrong leading me to conclude that the outlier is not the cause for this phenomenon. Since I expect a certain direction for the effects of all five regressors on female voter turnout a one-sided test is the appropriate mean to test the significance of the regressors. The illiteracy rate is the most insignificant regressor since a negative relationship is expected meaning the t-value of 2.87 should be smaller than the critical value of -1.72 in case of significance. For this reason I remove the illiteracy rate leading to model 3.[9] Homes owned is now the most insignificant regressor since a positive relationship is expected meaning its t-value of -0.68 should be larger than 1.72 in case of significance. Hence, I remove homes owned arriving at model 4. Now, voter turnout in 1912 is the most insignificant regressor since a positive relationship is expected meaning its t-value of -0.85 should be larger than 1.71 in case of significance. Removing voter turnout in 1912 leads to model 5. Here, the ratio of men to women is expected to have a negative relationship meaning its t-value of 0.18 should be smaller than the critical value of -1.71. Finally, only the ratio of foreign-born men survives the stepwise elimination of insignificant regressors. Even though a number of intuitive variables prove to be insignificant, the number of foreign-born men is a significant measure of the wealth of a ward since a high number of foreign-born men is a solid indicator of a large number of immigrants who typically were very poor.

Next, I repeat the analysis with the female voter participation rate, i.e., the percentage of female voters per ward relative to all voters. The main hypothesis is as before that richer wards have a higher number of women showing up at the polls. The results are similar. First, I look at the scatter diagrams on page 93. As before the diagrams depict relationships in line with the main hypothesis. The illiteracy rate, the ratio of men to women and the ratio of foreign-born men have a negative relationship and the number of homes owned and the voter turnout in 1912 have a positive relationship with the female voter participation rate. Also,

[9]Strictly speaking, I use the econometric method of stepwise regression. The idea is to start each step with eliminating the most insignificant regressor. Then the model is re-estimated in order to arrive at the next step. This procedure continues until the model does not contain any insignificant regressors anymore. Common criticisms of this procedure point to the fact that different orders of elimination lead to different final models. Also, the $F-$statistics and $t-$statistics are difficult to interpret. There are two reasons why I am using this unusual procedure here. First, I am choosing my model from a set of highly multicollinear variables. Second, I eliminate variables stepwise that are most insignificant with regard to a one-sided test and a theoretical prediction. In that sense, the $t-$statistics are fairly robust.

ward 2 is an outlier with respect to the illiteracy rate. Model 1 in table 3.5 on page 94 gives the regression results for all five regressors. The illiteracy rate, homes owned and the voter turnout in 1912 have the wrong sign. I check whether the outlier of ward 2 is causing this phenomenon. The results shown in the column of model 2 suggest that this is not the case as the signs for all regressors remain the same. As before, based on the expected sign of the regressors I conduct one-sided tests removing the most insignificant regressors step by step. Model 1 shows that the illiteracy rate is the most insignificant regressors because for it to be significant its t-value of 2.65 should be smaller than the critical value of -1.72. Removing the illiteracy rate leads to model 3. The voter turnout in 1912 is the most insignificant regressor since its t-value of -0.96 should be larger than the critical value of 1.72 for it to be significant. Removing the voter turnout of 1912 yields model 4. The most insignificant regressor is homes owned since its t-value of -1.29 needs to be larger than 1.71 for it to be significant. Removing homes owned leads to model 5. All regressors have the right sign. The ratio of men to women is the most insignificant regressor since its t-value of -0.02 needs to be smaller than -1.71 for it to be significant. Removing the ratio of men to women leaves foreign-born men as the only regressor which is also significant and has the right sign. Hence, I conclude that the number of foreign-born men is a significant determinant of female voter participation. With regard to the main hypothesis, this is evidence that a higher number of women are showing up at the polls in richer wards.

The second question of interest is whether the enactment of women suffrage had a significant impact on the election results. To this end, I look at the share of Republican votes in the presidential election of 1920. The question is whether the female voter turnout or the female participation rate had a significant impact on the share of Republican votes. First, I look at the scatter diagrams on page 95.

Obviously, the share of Republican votes in 1916 is highly positively correlated with the share of Republican votes in 1920. Ward 15 is somewhat of an outlier with only 40% of all votes being Republican. Female voter turnout and female participation rate seem to be positively correlated with the share of Republican votes. To my knowledge there is no strong party affiliation across wards. If there were one, one could hypothesize how this relates to the above result that more women are voting in richer wards. I also look at the change in illiteracy rate and the change in the ratio of men to women between the census years 1910 and 1920. The idea is to check whether a change in those variables might cause a change in the share of Republican votes. Both variables display a negative relationship. The change in the illiteracy rate has an outlier in ward 6 which had a significant improvement in the illiteracy rate. The ratio of foreign-born men displays a negative relationship while homes owned and voter turnout in 1912 have a positive relationship with the share of Republican votes. Again, there is little evidence to hypothesize about the directions of these relationships. In table 3.6 on page 96 I investigate whether female voter turnout is a significant determinant of the share of Republican votes. Model 1 includes all variables. Note that female voter turnout itself is a function of foreign-born men implying there might be an simultaneity issue. For this reason, I estimate the model without female voter turnout. The results are given in the column of model 2. All the insignificant factors are still insignificant leading me to conclude that the simultaneity issue is not too severe here. Since there are no hypotheses about the directions of the effects of the regressors I conduct two-sided test. An F-test with the null hypothesis that the change of the ratio of men to women, foreign-born men and homes owned are jointly insignificant does not reject the null at any conventional significance level. I remove these three regressors leading to model 3. The change in the illiteracy rate is insignificant. Removing the regressor leads to model 4. Now, female voter turnout is insignificant at the 5% significance level. Thus, I conclude that female voter turnout is not a significant determinant of the share of Republican votes in 1920.[10] I also check whether the female voter participation is a significant determinant of the share of Republican votes in 1920. The results are reported in table 3.7 on page 97. Again, there might be a simultaneity issue since female voter participation depends on the ratio of foreign-born men. Comparing the results of model 2 with the results of model 1 shows that the insignificant variables remain insignificant leading me to conclude that the simultaneity issue is not too severe here. I follow the same procedure as before removing the most insignificant regressors step by step. This leads me to the same conclusion as before that female voter participation is not a significant determinant of the share of Republican votes.

Table 3.8 on page 98 offers an interesting background for these results. The last three columns show that the share of Republican votes fell from 1912 to 1916 and rebounded again in 1920. Many wards show this V-shaped pattern. This suggests that there was maybe some other national factor at place which was more important than the enactment of women suffrage. Also, it needs to be kept in mind that the so-called (Republican) machine in Pittsburgh was still very powerful implying that the local Republicans might have been able to offset any possible effect of women suffrage by making sure that the additional female voter turnout in the richer wards would still vote Republican.

To conclude, I find that there was higher female voter turnout and a higher female voter participation rate in richer wards while I could not find any effects of the enactment of women suffrage on the election outcome.

[10] Note that my sample is very small. It contains only 27 observations. This makes it difficult to argue that the same insignificant result would be found in a larger sample. In fact, a larger sample might lead to a significant result.

3.7 The Analysis with Voter Registration Data

The above analysis is complemented with information on voter registration. I have access to the lists of registered voters per ward and election district for various years provided by the Pittsburgh Board of Registration Committee Records (1906–1936). I can see the name of the registered voter but not his or her party affiliation. The simple idea is to use the first name of the registered voter to identify the gender in order to learn about female voter participation rates.

Since it was not possible to record the gender for every registered voter I recorded the gender per voter for the first 200 registered voters per ward. The procedure is as follows. Every ward consists of a number of election districts, say n. In order to cover the whole ward I record the gender of the first $200/n$ registered voters for each election district if the particular ward in case has n election districts. That means that I record between 6 and 50 voters per election district. Within each election district the voters are ordered per street and house number. Then, I simply start at the top recording the gender since how the streets or voters per streets are ordered does not seem to follow a particular pattern. However, there might be an issue of sample selection here. It is possible that the first entries of each election district might not be random but are chosen in some way by the person putting together the registration lists. For example, the lists might be structured in a way that the first entries of each street contain houses where a particularly high number of women lived. Also, picking the same number of voters across election districts within a particular ward might give too much weight to small districts which might have a particular gender structure. Another problem is that for some names a gender could not be decided. A typical example is the use of initials in the voter lists. In that case, I simply move to the next voter. A natural concern is if there is a gender-specific bias in the sense that unidentifiable names rather tend to belong to a specific gender. The correlation coefficient between the percentage of female voters per ward and the number of unidentifiable voters is 0.14. That means there is a slight bias in the sense that a high number of unidentifiable voters goes hand in hand with a high number of female voters.

Table 3.9 on page 99 displays the percentage of female registered voters among all recorded registered voters in columns three and four. I have two incomplete group of data; one for 1925 and one for the early 1930s. In 1925, data for wards $19-24$ are missing. Also, for the early 1930s I only have data for a period of years. The data for wards $1-8$ are from 1931, for wards $9-16$ are from 1932, for wards $17-21$ are from 1933 and for wards $25-32$ are from 1931. Data for wards $22-24$ are missing for the early 1930s. Also, note that in 1920 Pittsburgh had 27 wards, in 1924 a 28th ward was added, in 1928 wards 29 and 30 while finally in 1932 wards 31 and 32 were added.

Because of the sample selection issues, the bias towards female names, the missing data problem and the fact that I do not have registration data for the initial election after the enactment of women suffrage, I will discuss these data only informally in the context of the previous results. Table 3.9 on page 99 shows the female participation rate for 1920 based on the interpolation exercise in the previous section as well as the female participation rate based on the voter registration data. The estimates of the interpolation exercise show that the wards 4, 16, 17 and 24 have the lowest female participation rate in 1920. The map on page 100 shows that all of these wards are in the flood plains near the rivers suggesting these are predominantly working class wards. On the other hand, the wards 7,8,11 and 14 are mainly up the hills in the East End of Pittsburgh suggesting these are middle and upper class wards. This confirms the hypothesis that female voter participation is lower in working class neighborhoods. The voter registration for 1925 show that wards 2, 16, 21 and 24 had the lowest female voter participation rates. Again, all of these wards are close to the rivers in the flood plains. The wards with the highest female voter participation rates are 7,8,25 and 28. The maps on pages 100 and 101 show that neither of these wards has access to the rivers confirming the previous finding that the wards with high female voter participation rates are not working class wards. Finally, the data for the early 1930s show that the wards 2,3,12 and 18 have the lowest female participation rates. Here the picture is ambiguous. The wards 2 and 3 are clearly close to the rivers and the mills while ward 12 is in the East End and ward 18 is mainly outside the flood plains. The picture is clearer for the wards with a high female participation rate. Even though ward 27 is near the Ohio river, wards 7,8,28 and 32 are clearly far away from the flood plains near the rivers.

However, the data for the early 1930s should be interpreted with caution. First, these data are from a decade after the enactment of women suffrage. The analysis in this paper on the other hand, is concerned with the immediate effect of enacting women suffrage. In other words, a lot of things might have happened within ten years making it difficult to use the data from the early 1930s to interpret the initial effect of women suffrage in 1920. Second, the Great Depression as well as Roosevelt's New Deal had changed the political landscape dramatically making it a challenging choice for the study of female voting behavior. Nevertheless, I take at least the voter registration data for 1925 as a confirmation of the earlier exercise, i.e., the female voter participation rates were lower in working class wards.

The data for the early 1930s however, are interesting in another respect. 18 wards have an estimate for female voter participation for 1920, 1925 and the early 1930s. 14 of those have a strong upward trend. For example, the time path for the working class ward 1 reads 11%, 31% and 39% while it reads 40%, 42% and 48% for the upper class ward 8. Even though these are voter participation data they seem to suggest strongly that over time more and more women did show up at the polls. Similarly, Lott and Kenny (1999) show that female voter turnout increases over time after the initial enactment of women suffrage. This is in line with the general hypothesis that it takes time for a newly enfranchised group to show up at the polls after the initial extension of voting rights.

3.8 Conclusion

This work is concerned with the effect of extending the franchise to women. The motivating question in the background is why women had to fight for their right to vote. The hypothesis is that extending the franchise to women has different consequences for different social groups. In particular, the idea is that richer and more educated women show up in larger numbers at the polls than poorer and less-educated women. If this were true this might imply that the richer and more educated groups in society are proponents of extending the franchise to women while the poorer and less-educated groups in society are opposing this idea. I study this question for the case of Pittsburgh in 1920 when women first had the right to vote in Pennsylvania.

I have three sets of results. First, I interpolate the number of male voters in the presidential election of 1920 from the presidential election of 1916 and back out the number of female voters per ward. Qualitative evidence on the location of poor and rich neighborhoods in Pittsburgh at that time suggests that female voter turnout and female voter participation were higher in richer and more educated neighborhoods. Second, I regress female voter turnout and female voter participation on a number of variables indicating social status. I find that the number of foreign-born men has a negative effect on female voter turnout and female voter participation. Male immigrants often came first alone before bringing their wives over. Since these immigrants were often poor and had left school early I conclude that female voter turnout and female voter participation was higher in the richer parts of town. The next question is whether the enactment of women suffrage had a significant effect on the election results. To this end, I regress the share of Republican votes in 1920 on female voter turnout and female voter participation. I cannot find a significant effect of these variables on the share of Republican votes. There might be a number of reasons why there is no effect. There might have been election-specific issues which dominated the effect of the enactment of women suffrage. On the other hand, the influence of the local Republican party in Pittsburgh might have been too dominant to allow for a significant effect of women suffrage. Third, I look at voter registration data from 1925 and the early 1930s. Qualitative evidence on the location of poor and rich neighborhoods suggests that more female voters registered in rich neighborhoods than in poor neighborhoods in 1925. Also, comparing the female voter participation rates from 1920, 1925 and the early 1930s suggest that female voter participation was rising after the initial enactment of women suffrage confirming the hypothesis that it takes some time for a newly enfranchised group to show up at the polls. I view this work as a pilot study. The reason is that I only discuss the voter registration data qualitatively. Future work would record the voter registration data much more carefully, so that these data could be used for statistical and econometric inference.

The overall conclusion is that there is some evidence for the case of Pittsburgh suggesting that richer and more educated groups might have had a reason to support women suffrage while poorer and less educated groups had a reason to oppose women suffrage. This line of research can be extended in a number of ways. First, working with voter registration data showing party affiliation might be an advantage. Second, controlling for the influence of the local Republican party on election results might improve the results. Third, in order to make conclusions beyond a case study it is necessary to tackle the sample selection issue intelligently since it will be impossible to record all voter registration records for 1920 nationally.

3.9 Appendix

The following quotes are from the historic novel *Out of this Furnace* by Bell (1950). They show the influence the large steel mills in early twentieth century Pittsburgh had on the voting behavior of their workers. In the quote below the term "local option" refers to the idea that the legal status of selling alcohol should be decided on a local level.

> That was an election year, and the company spread word that if the Republicans won, and more particularly the Penrose men – local option had become a thread to the machine – the mills might resume full-time operations. It did more. During the May primaries a number of departmental bosses prominent in the local option movement were called into the General Superintendent's office and ordered to support the Penrose candidates, who were of course identified with the liquor interests. They were told their first duty was to the company and the company needed Penrose in the Senate and, therefore, Penrose men in the State Legislature; popular election of Senators was still five years in the future. And the summoned ones obeyed, bosses though they were, while from Duquesne came the usual reports of men fired for refusing to vote as the company wished.[11]

> On election day Hooker, the head blower, came into the casthouse and said he'd take over while Keogh went to vote. "How many men you got?" "Five or Six." "Round them up." Mike proved to be the only Slovak in the group. Hooker stared at him. "You a citizen?" he asked. He always did. "Yessir. Pretty soon ten years." "Humph. Well, I just

[11] Page 160.

want to remind you that it's the company that pays your wages, not some politician in Harrisburg or Washington. Anything that hurts the company hurts you, remember that. And the company hasn't got much use for a man that can't be loyal. Just keep that in mind when you vote. Take these along; it's all right to take them into the booth with you." He passed out sample ballots marked with a vote for a straight Republican ticket.[12]

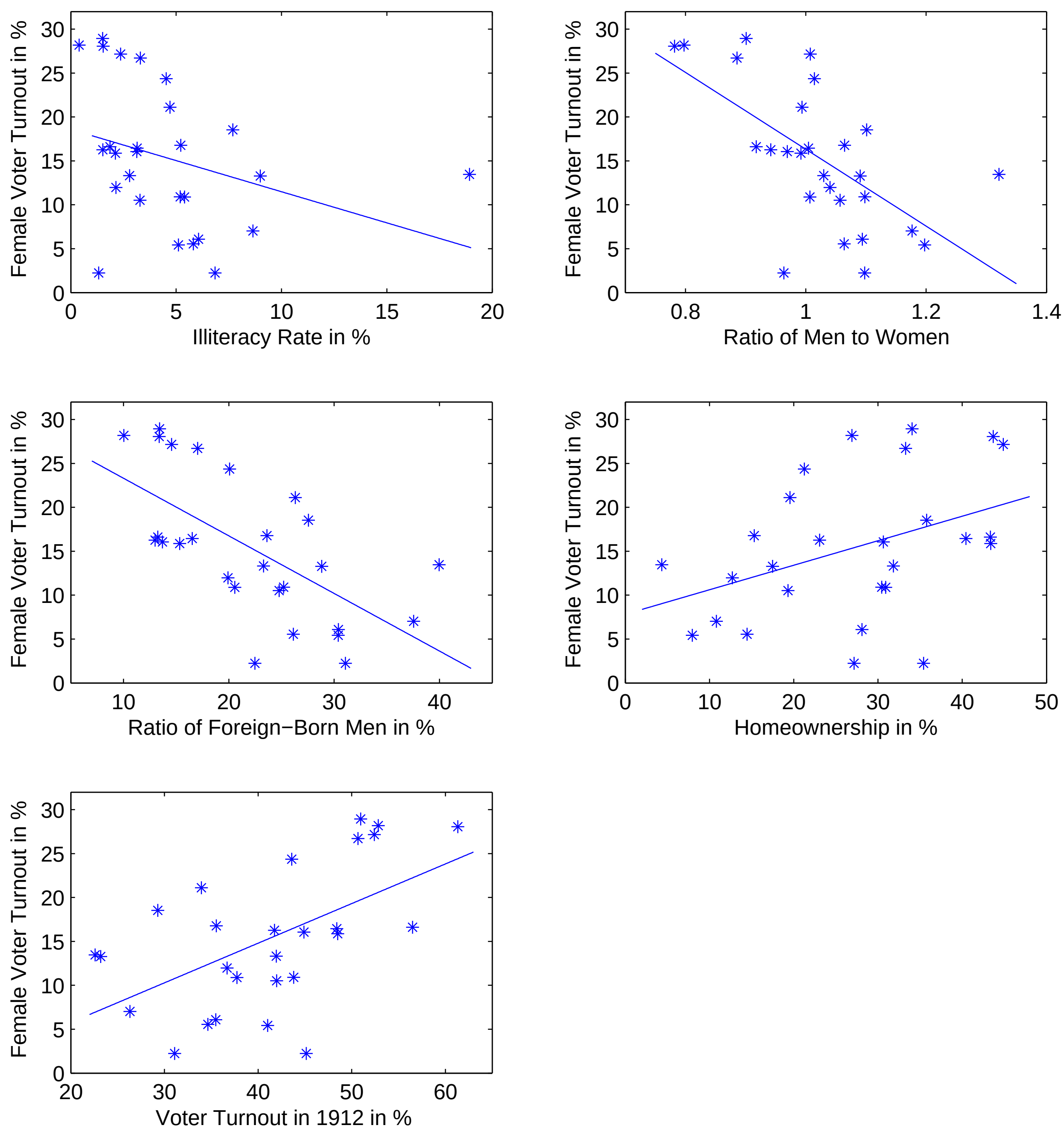
[12] Page 187.

Ward	Male Turnout 1916	Female Turnout 1920	Female Participation Rate 1920
1	34.7	5.4	10.6
2	31.9	13.5	21.2
3	27.3	7.0	16.7
4	37.5	21.1	35.8
5	41.2	24.4	35.8
6	36.7	13.3	23.9
7	48.4	28.2	43.4
8	48.4	26.7	39.5
9	43.8	10.5	17.8
10	43.1	13.3	22.7
11	47.9	28.9	40.4
12	42.3	10.9	19.9
13	46.3	16.1	26.5
14	57.8	28.1	39.3
15	34.8	10.9	20.8
16	35.8	2.2	4.9
17	36.2	6.1	12.7
18	49.9	15.9	24.6
19	50.3	27.2	35.5
20	51.2	16.5	23.9
21	37.6	16.8	29.2
22	32.2	12.0	26.1
23	34.5	5.6	12.6
24	46.2	2.2	4.7
25	42.8	16.3	29.1
26	49.6	16.6	27.4
27	35.4	18.6	30.9

Female turnout is the percentage of all men older than 21 voting relative to all male voters.
Female participation rate is the percentage of the female voters relative to all voters.

Table 3.3. Estimation Results for Female Voter Turnout in 1920.

Figure 3.2. Scatter Diagrams of Female Voter Turnout in 1920.

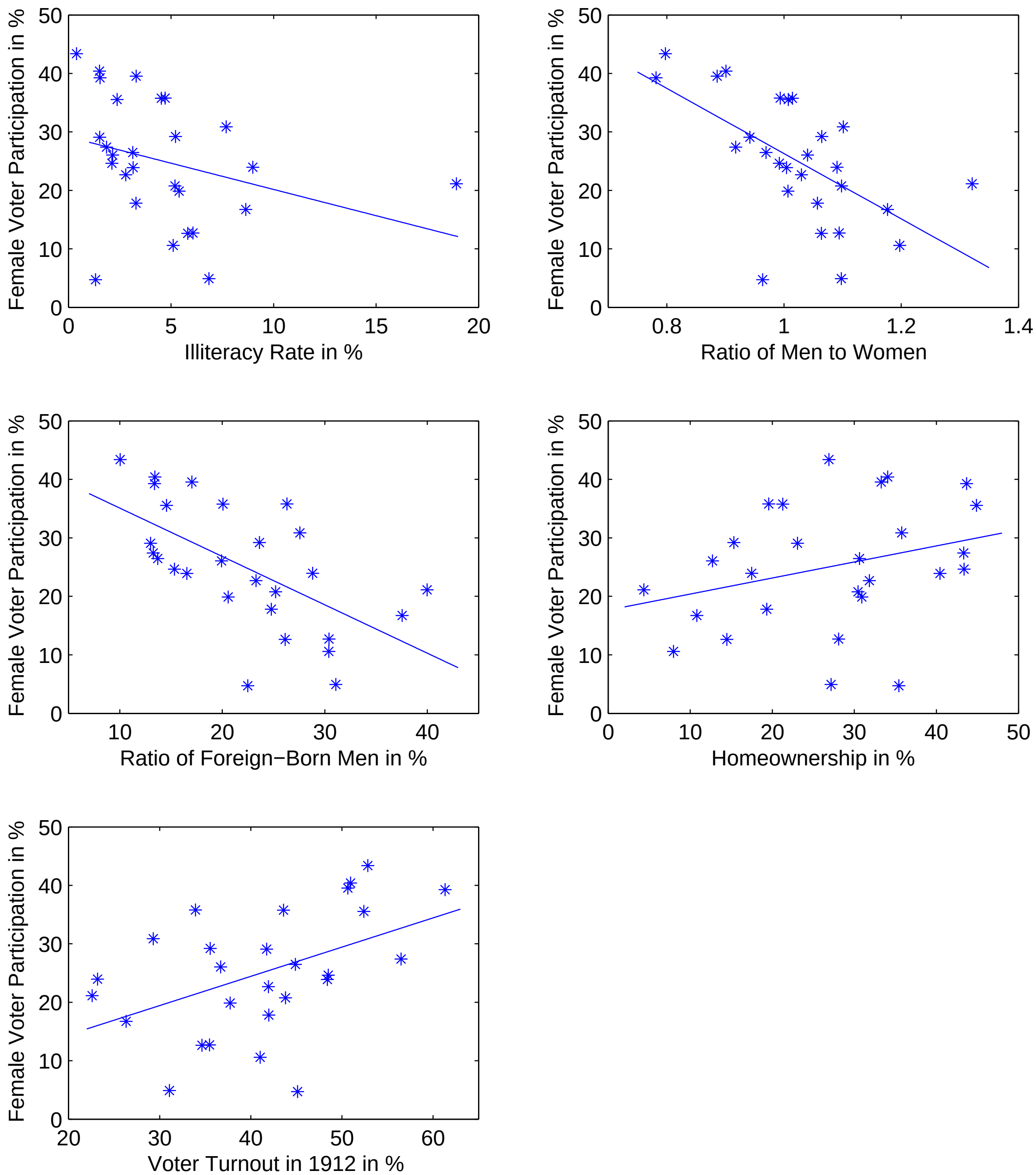


Female Voter Turnout	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	-1.87	-1.92	-4.06	-3.89	-3.34	-3.22
	(-1.56)	(-1.59)	(-3.8)	(-3.79)	(-4.19)	(-8.55)
Illiteracy Rate	0.90	0.83				
	(2.87)	(2.53)				
Ratio of Men to Women	-3.05	-3.2	-0.41	-0.08	0.41	
	(-1.29)	(-1.34)	(-0.16)	(-0.03)	(0.18)	
Foreign-Born Men	-2.02	-1.92	-1.4	-1.38	-1.1	-1.01
	(-3.25)	(-3.01)	(-2.08)	(-2.08)	(-1.92)	(-3.74)
Homes Owned	-0.35	-0.26	-0.18			
	(-1.48)	(-0.98)	(-0.68)			
Voter Turnout 1912	0.07	0.02	-0.38	-0.5		
	(0.13)	(0.04)	(-0.61)	(-0.85)		
Dummy Ward 2		0.61				
		(0.8)				
Adjusted R^2	0.46	0.45	0.28	0.3	0.31	0.33
Schwarz	2.23	2.32	2.44	2.34	2.25	2.12
Hannan-Quinn	2.02	2.08	2.27	2.2	2.14	2.06
Critical Value	1.72	1.73	1.72	1.71	1.71	1.71

Female Voter Turnout is the independent variable.
T-values in parenthesis.
Critical value is for a one-sided t-test at the 5% significance level.

Table 3.4. OLS Estimation Results for Female Voter Turnout in 1920.

Figure 3.3. Scatter Diagrams of Female Voter Participation in 1920.

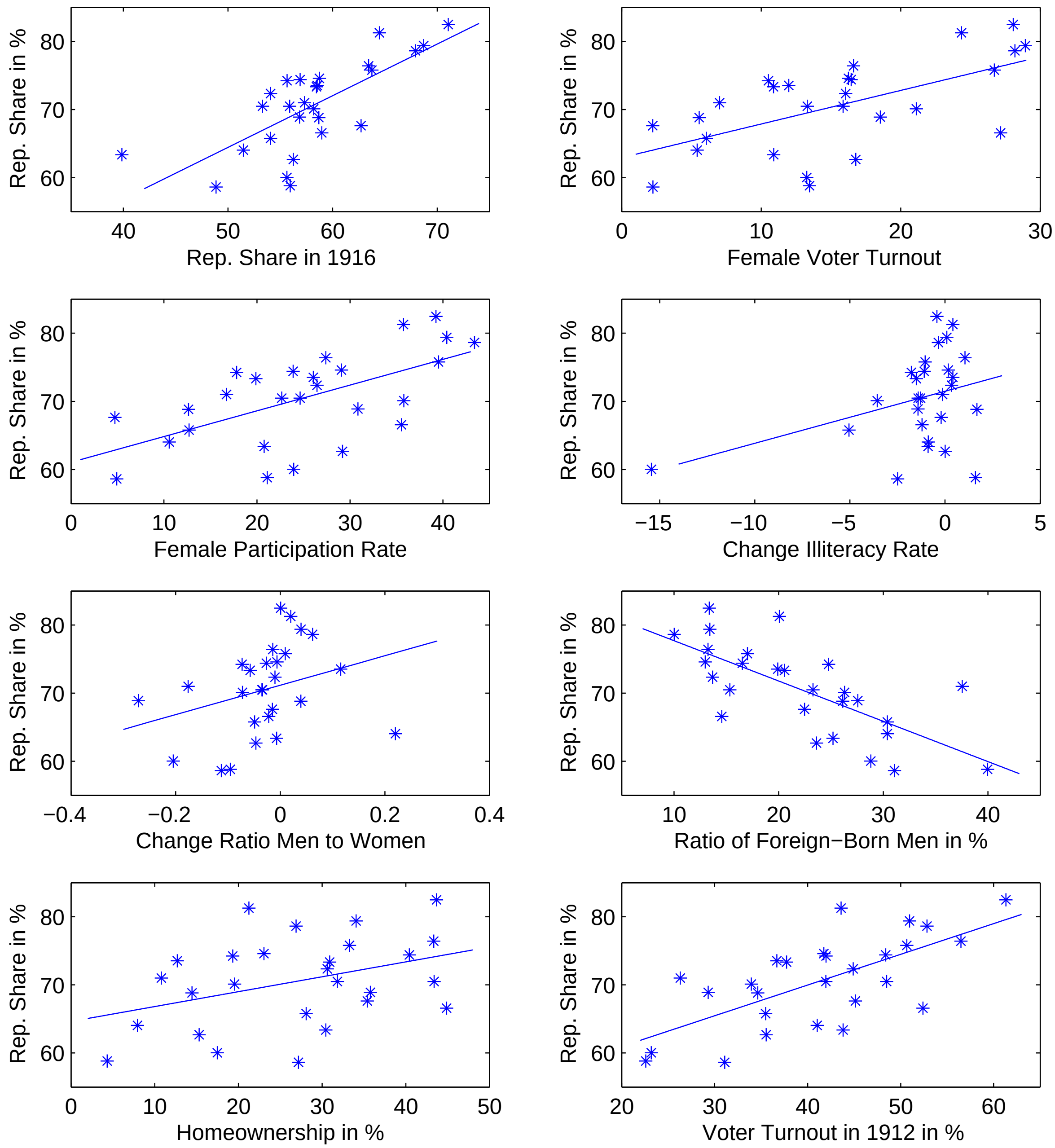


Female Voter Participation Rate	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	-1.46	-1.48	-3.3	-2.83	-2.33	-2.35
	(-1.33)	(-1.31)	(-3.46)	(-3.45)	(-3.18)	(-6.79)
Illiteracy Rate	0.76	0.73				
	(2.65)	(2.4)				
Ratio of Men to Women	-3.35	-3.4	-1.13	-0.77	-0.05	
	(-1.55)	(-1.54)	(-0.5)	(-0.35)	(-0.02)	
Foreign-Born Men	-1.78	-1.75	-1.26	-0.99	-0.86	-0.86
	(-3.14)	(-2.96)	(-2.1)	(-1.87)	(-1.62)	(-3.48)
Homes Owned	-0.36	-0.33	-0.22	-0.29		
	(-1.69)	(-1.36)	(-0.94)	(-1.29)		
Voter Turnout 1912	-0.16	-0.17	-0.53			
	(-0.3)	(-0.33)	(-0.96)			
Dummy Ward 2		0.23				
		(0.32)				
Adjusted R^2	0.44	0.41	0.29	0.29	0.27	0.3
Schwarz	2.04	2.16	2.21	2.13	2.07	1.95
Hannan-Quinn	1.84	1.93	2.04	1.99	1.97	1.88
Critical Value	1.72	1.73	1.72	1.71	1.71	1.71

Female Voter Participation Rate is the independent variable.
T-values in parenthesis.
Critical value is for a one-sided t-test at the 5% significance level.

Table 3.5. OLS Estimation Results for Female Voter Participation Rate in 1920.

Figure 3.4. Scatter Diagrams of the Share of Republican Votes in 1920.



Share of Republican Votes 1920	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	0.94	0.57	0.98	0.94	0.79
	(2.07)[0.05]	(1.92)[0.07]	(6.84)[0.0]	(6.76)[0.0]	(9.5)[0.0]
Share of Republican Votes 1916	0.57	0.59	0.57	0.59	0.67
	(3.0)[0.007]	(3.08)[0.006]	(3.45)[0.002]	(3.62)[0.002]	(4.35)[0.0002]
Female Voter Turnout 1920	0.08		0.08	0.07	
	(1.08)[0.3]		(1.49)[0.15]	(9.5)[0.2]	
Change Illiteracy Rate 1910-20	0.11	0.08*	0.1		
	(0.97)[0.35]	(0.74)[0.46]	(1.08)[0.29]		
Change Ratio Men to Women 1910-20	0.03*	-0.48*			
	(0.03)[0.98]	(-0.52)[0.61]			
Foreign-Born Men 1920	-0.02*	-0.14*			
	(-0.1)[0.92]	(-0.8)[0.43]			
Homes Owned 1920	0.01*	-0.06*			
	(0.08)[0.94]	(-0.47)[0.65]			
Voter Turnout 1912	0.23	0.33	0.27	0.28	0.32
	(0.8)[0.43]	(1.17)[0.26]	(2.7)[0.01]	(2.89)[0.01]	(3.28)[0.003]
Adjusted R^2	0.64	0.64	0.69	0.69	0.67
Schwarz	0.19	0.12	-0.18	-0.25	-0.3
Hannan-Quinn	-0.08	-0.11	-0.35	-0.38	-0.4
F-Test	0.01 [0.99]	(0.35)[0.83]			

Share of Republican Votes 1920 is the independent variable.
T-values in parenthesis, p-values in brackets.
F-test is a Wald test with the null that all coefficients marked with an asterisk are jointly insignificant.

Table 3.6. OLS Estimation Results for a Shift in Election Results due to Female Voter Turnout.

Share of Republican Votes 1920	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	1.01	0.57	0.96	0.93	0.79
	(2.38)[0.03]	(1.92)[0.07]	(8.1)[0.0]	(8.03)[0.0]	(9.5)[0.0]
Share of Republican Votes 1916	0.56	0.59	0.55	0.16	0.67
	(3.0)[0.007]	(3.08)[0.006]	(3.4)[0.003]	(3.54)[0.002]	(4.35)[0.0002]
Female Voter Participation 1920	0.1		0.1	0.1	
	(1.42)[0.17]		(1.8)[0.09]	(1.68)[0.11]	
Change Illiteracy Rate 1910-20	0.11	0.08*	0.1		
	(1.01)[0.32]	(0.74)[0.46]	(1.08)[0.29]		
Change Ratio Men to Women 1910-20	0.1*	-0.48*			
	(0.11)[0.92]	(-0.52)[0.61]			
Foreign-Born Men 1920	0.02*	-0.14*			
	(0.1)[0.92]	(-0.8)[0.43]			
Homes Owned 1920	0.03*	-0.06*			
	(0.19)[0.85]	(-0.47)[0.65]			
Voter Turnout 1912	0.24	0.33	0.28	0.29	0.32
	(0.8)[0.4]	(1.17)[0.26]	(2.87)[0.009]	(3.05)[0.01]	(3.28)[0.003]
Adjusted R^2	0.65	0.64	0.7	0.7	0.67
Schwarz	0.15	0.12	-0.21	-0.29	-0.3
Hannan-Quinn	-0.13	-0.11	-0.39	-0.42	-0.4
F-Test	0.02 [0.99]	(0.35)[0.83]			

Share of Republican Votes 1920 is the independent variable.
T-values in parenthesis, p-values in brackets.
F-test is a Wald test with the null that all coefficients marked with an asterisk are jointly insignificant.

Table 3.7. OLS Estimation Results for a Shift in Election Results due to Female Voter Participation.

Ward	Republican 1912	Republican and Roosevelt 1912	Republican 1916	Republican 1920
1	34	67	51	64
2	49	72	56	59
3	34	68	57	71
4	26	64	58	70
5	19	67	64	81
6	31	60	56	60
7	31	67	68	79
8	24	61	64	76
9	29	61	56	74
10	19	55	53	70
11	23	67	69	79
12	18	59	58	73
13	13	56	54	72
14	30	70	71	82
15	23	50	40	63
16	16	51	49	59
17	21	53	54	66
18	15	57	56	70
19	15	61	59	67
20	27	63	57	75
21	18	58	56	63
22	24	66	59	74
23	22	61	59	69
24	20	51	63	68
25	12	63	59	75
26	13	62	64	76
27	23	61	57	69

All numbers in %.
Roosevelt lost the Republican convention in 1912 and ran on his own ticket.

Table 3.8. Share of Republican Votes in Presidential Elections in Pittsburgh.

Ward	1920*	1925**	1930s**
1	11	31	39
2	21	28	29
3	17	33	33
4	36	30	37
5	36	29	44
6	24	31	41
7	43	44	54
8	40	42	48
9	18		37
10	23		40
11	40		47
12	20		36
13	27		40
14	39		42
15	21	31	36
16	5	25	40
17	13	30	36
18	25	31	36
19	36	39	47
20	24	37	46
21	29	28	44
22	26	39	
23	13	31	
24	5	20	
25	29	39	44
26	27	32	44
27	31	29	48
28		44	48
29			42
30			44
31			47
32			53

All numbers in %.
* Previous interpolation.
** Registration records.

Table 3.9. Percentage of Female Voter Participation Rate.

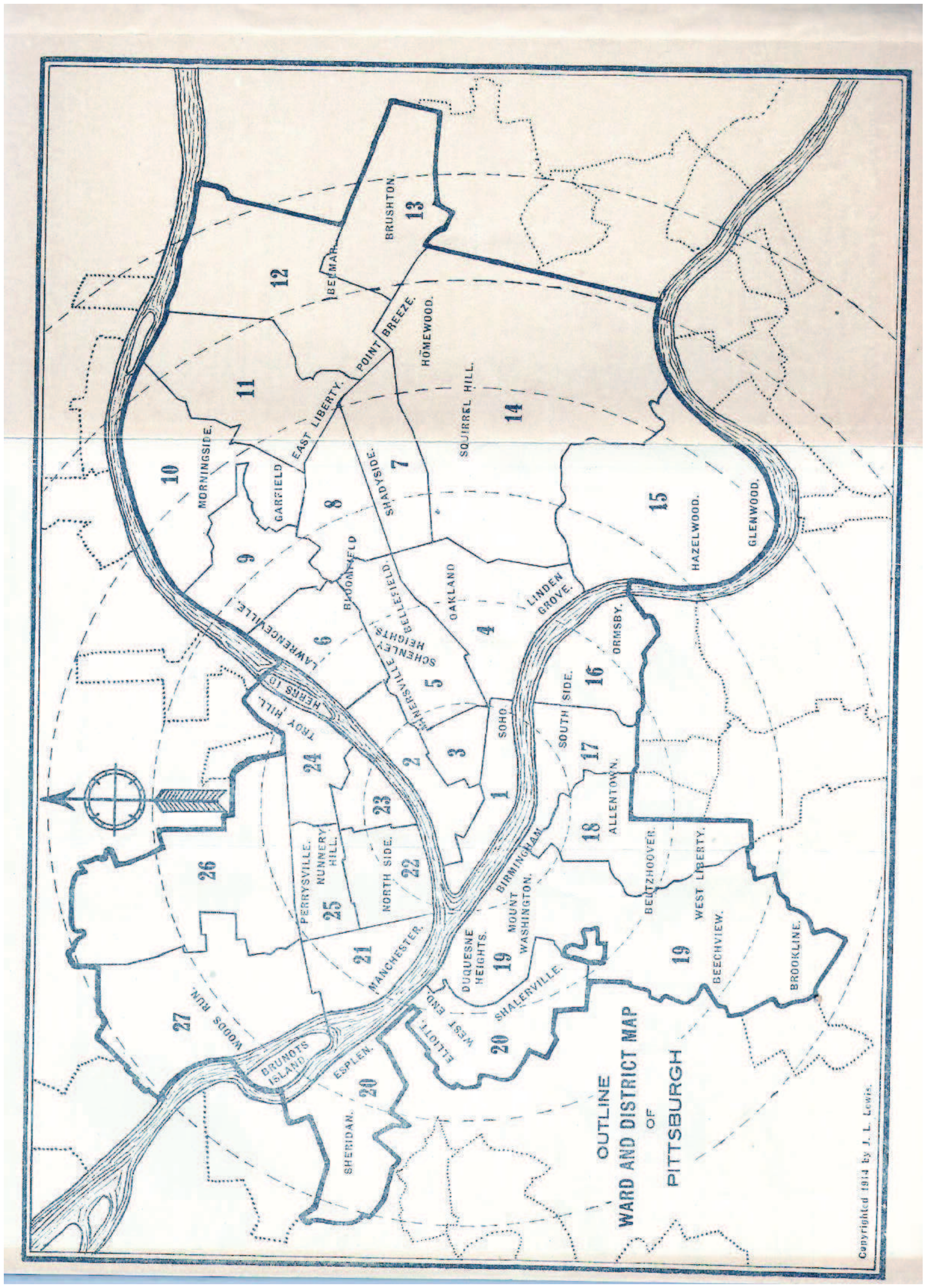


Figure 3.5. The 27 Wards of Pittsburgh in 1914.

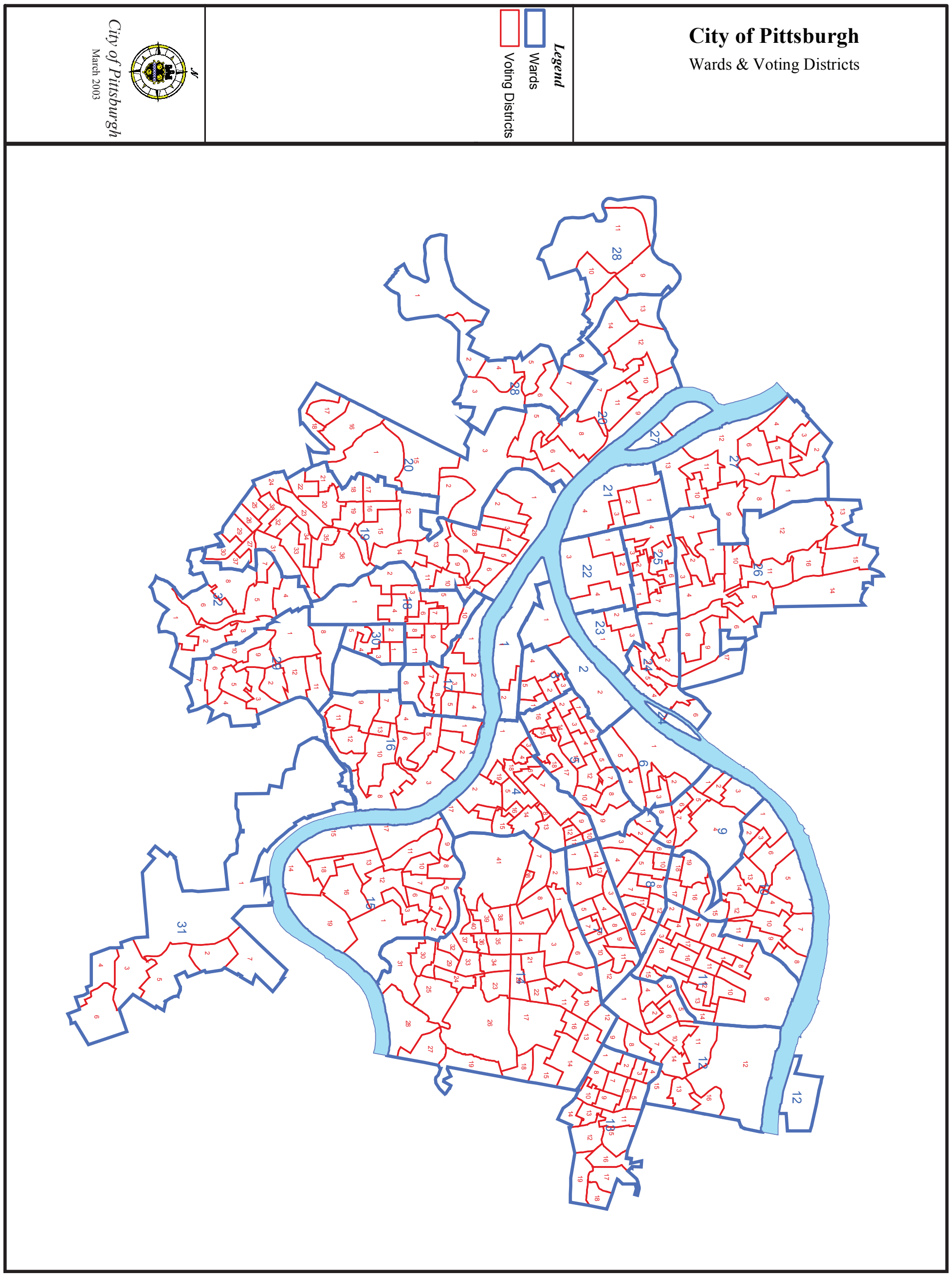


Figure 3.6. The 32 Wards of Pittsburgh in 2003.

Bibliography

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward A Theory Of Discounted Repeated Games With Imperfected Monitoring," *Econometrica*, 58(5), 1041–1063.

Acemoglu, D., S. Johnson, J. Robinson, and P. Yared (2008): "Income and Democracy," *American Economic Review*, 98, 808–842.

Acemoglu, D., and J. Robinson (2000): "Why did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective," *Quarterly Journal of Economcis*, 115(4), 1167–1199.

Alesina, A., and E. Spolaore (1997): "On the Number and Size of Nations," *The Quarterly Journal of Economics*, 112(4), 1027–1056.

Alvarez, F., and U. Jermann (2001): "Quantitative Asset Pricing Implications of Endogeneous Solvency Constraints," *Review of Financial Studies*, 14(4), 1117–1151.

Alvarez, M., J. Cheibub, F. Limongi, and A. Przeworski (1996): "Classifying Political Regimes," *Comparative International Development*, 31(2), 3–36.

Attanasio, O., L. Picci, and A. Scorcu (2000): "Saving, Growth, and Investment: A Macroeconomic Analysis Using a Panel of Countries," *The Review of Economics and Statistics*, 82(2), 182–211.

Barrett, J., and D. Roediger (1997): "Race, Nationality and the "New Immigrant" Working Class," *Journal of American Ethnic History*, 16(3), 3–44.

Barro, R. (1999): "Determinants of Democracy," *Journal of Political Economy*, 107(6), 158–183.

Bell, T. (1950): *Out of this Furnace*. University of Pittsburgh Press.

Berman, D. (1993): "Gender and Issue Voting: The Policy Effects of Suffrage Expansion in Arizona," *Social Science Quarterly*, 4, 838–850.

Bierens, H. (2009): *Specification of Econometric Models*. Unpublished Manuscript, Pennsylvania State University.

Bodnar, J., S. Roger, and M. Weber (1982): *Lives of Their Own: Blacks, Italian, and Poles in Pittsburgh, 1900–1960*. University of Illinois Press.

Bollen, K. (1990): "Political Democracy: Conceptual and Measurement Traps," *Studies in Comparative International Development*, 25, 7–24.

Bollen, K., and R. Jackman (1989): "Democracies, Stabilities and Dichotomies," *American Sociological Review*, 54, 612–621.

Casper, G., and C. Tufis (2003): "Correlation versus Interchangeability: The Limited Robustness of Empirical Findings on Democracy using Highly Correlated Datasets," *Political Analysis*, 11(2), 196–203.

Census (1910): *Thirteenth Census of the United States Taken in the Year 1910–Statistics for Pennsylvania*. Washington, Government Printing Office.

——— (1920): *Fourteenth Census of the United States–State Compendium Pennsylvania*. Washington, Government Printing Office.

Corder, K., and C. Wolbrecht (2006): "Political Context and the Turnout of New Women Voters after Suffrage," *The Journal of Politics*, 68(1), 34–49.

Dahl, R. (1971): *Polyarchy*. Yale University Press.

Elkins, Z. (2000): "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations," *American Journal of Political Science*, 44(2), 287–294.

Epstein, D., R. Bates, J. Goldstone, I. Kristensen, and S. O'Halloran (2006): "Democratic Transitions," *American Journal of Political Science*, 50(3), 551–569.

Feller, W. (1950): *An Introduction to Probability Theory and its Applications*. Wiley.

Flexner, E., and E. Fitzpatrick (1996): *Century of Struggle: The Woman's Rights Movement in the United States*. Harvard University Press.

Freedom House (2008): *Map of Freedom 2008*. http://www.freedomhouse.org /uploads/MOF2008.pdf (04/13/09).

——— (2009): *Freedom in the World*. http://www.freedomhouse.org /uploads/fiw/FIWAllScores.xls (03/31/09).

Fudenberg, D., and J. Tirole (1991): *Game Theory*. MIT Press.

Funk, P., and C. Gathmann (2006): "What Women Want: Suffrage, Gender Gaps in Voter Preferences and Government Expenditures," Unpublished Manuscript.

Gleditsch, K., and M. Ward (2001): "Measuring Space: A Minimum Distance Database," *Journal of Peace Research*, 38, 749–768.

Goldstein, J. (1984): *The Effects of the Adoption of Womens Suffrage: Sex Differences in Voting Behavior–Illinois, 1914-21*. New York: Praeger.

Granger, C. (1969): "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, 37(3), 424–438.

Granger, C., and P. Newbold (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

Gundlach, E., and M. Paldam (2008): *Income and Democracy: A comment on Acemoglu, Johnson, Robinson and Yared (2008)*. Kiel Working Paper No. 1458, Kiel Institute for the World Economy.

Heston, A., R. Summers, and B. Aten (2006): *Penn World Table Version 6.2*. Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Huntington, S. (1991): *The Third Wave: Democratization in the Late Twentieth Century*. Norman: Univ. Oklahoma Press.

Hurlin, C., and B. Venet (2001): *Granger Causality Tests in Panel Data Models with Fixed Coefficients*. Working Paper. EURIsCO, Universite Paris IX Dauphin.

Jaeger, D., and D. Paserman (2008): "The Cycle of Violence? An Empirical Analysis of Fatalities in the Palestinian-Israeli Conflict," *American Economic Review*, 98(4), 1591–1604.

Kocherlakota, N. (1996): "Implications of Efficient Risk Sharing without Commitment," *Review of Economic Studies*, 63, 595–609.

Koppes, C., and W. Norris (1985): "Ethnicity, Class, and Morality in the Industrial City: A Case Study of Typhoid Fever in Pittsburgh, 1890–1910," *Journal of Urban History*, 11(3), 259–279.

Krueger, D., and F. Perri (2006): "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *Review of Economic Studies*, 73, 163–193.

Leamer, E. (1985): "Vector Autoregressions for Causal Inference?," *Carnegie-Rochester Conference Series on Public Policy*, 22, 255–304.

Lipset, S. (1959): "Some Social Requisites of Democracy: Economic Development and Political Legitimcay," *American Political Science Review*, 53, 69–105.

Lott, J., and L. Kenny (1999): "Did Women's Suffrage Change the Size and Scope of Government?," *Journal of Political Economy*, 107(6), 1163–1198.

Luconi, S. (1996): "Machine Politics and the Consolidation of the Roosevelt Majority: The Case of Italian Americans in Pittsburgh and Philadelphia," *Journal of American Ethnic History*, 15(2), 32–59.

Mailath, G., and L. Samuelson (2006): *Repeated Games and Reputations*. Oxford University Press.

Marshall, M., and K. Jaggers (2004): *Political Regime Charactersitics and Transistions, 1800-2002*. Polity IV Project, University of Maryland.

Mayfield, L. (1993): "Voting Fraud in Early Twentieth-Century Pittsburgh," *Journal of Interdisciplinary History*, 24(1), 59–84.

Miller, G. (2008): "Womens Suffrage, Political Responsiveness, and Child Survival in American History," *Quarterly Journal of Economcis*, 123(3), 1287–1327.

Mosconi, R., and R. Seri (2006): "Non-causality in bivariate binary time series," *Journal of Econometrics*, 132, 379–407.

Muller, E. (2001): "Industrial Suburbs and the Growth of Metropolitan Pittsburgh, 1870–1920," *Journal of Historical Geography*, 27(1), 58–73.

Nelson, C., and C. Plosser (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139–162.

Ogburn, W., and I. Goltra (1919): "How Women Vote," *Political Science Quarterly*, 34, 413–433.

Pittsburgh Board of Registration Committee Records (1906–1936):. AIS.1983.19, Archives Service Center, University of Pittsburgh.

Przeworski, A., M. Alvarez, J. Cheibub, and F. Limongi (2000): *Democracy and Development*. Cambridge University Press.

Robinson, J. (2006): "Economic Development and Democracy," *Annual Reviews of Political Science*, 9, 503–527.

Roessing, J. B. (1914): "The Equal Suffrage Campaign in Pennsylvania," *Annals of the American Academy of Political and Social Science*, 56, 153–160.

Ronning, G., and M. Kukuk (1996): "Efficient Estimation of Ordered Probit Models," *Journal of the American Statistical Association*, 91(435), 1120–1129.

Sims, C. (1972): "Money, Income and Causality," *American Economic Review*, 62(4), 540–552.

——— (1980): "Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered," *American Economic Review*, 70, 250–257.

Smull, J. A. (1897–1949): *Smull's legislative hand book and manual of the state of Pennsylvania*. Harrisburg, Pennsylvania.

Stock, J., and M. Watson (2001): "Vector Autoregressions," *Journal of Economic Perspectives*, 15(4), 101–115.

Tertilt, M., and M. Doepke (2009): "Women's Liberation: What's in it for Men?," *Quarterly Journal of Economcis*, 124(4), 1541–1591.

Thomas, J., and T. Worrall (1988): "Self-Enforcing Wage-Contracts," *Review of Economic Studies*, 55, 541–554.

——— (1994): "Foreign Direct Investment and the Risk of Expropriation," *Review of Economic Studies*, 61, 81–108.

Treier, S., and R. Jackman (2008): "Democracy as a Latent Variable," *American Journal of Political Science*, 52(1), 201–217.

Valentino, B., P. Huth, and D. Balch-Lindsay (2004): "Draining the Sea: Mass Killings and Guerilla Warfare," *International Organization*, 58(2), 375–407.

Van Damme, E. (1987): *Stability and Perfection of Nash Equilibria*. Springer.

Vita

Sascha Claudius

Sascha Claudius was born on August 14, 1978, in Berlin, Germany. He earned his M.A. degree in Philosophy and Economics in 2005 from the Humboldt-University Berlin, Germany. Since then, he has continued his studies at the Pennsylvania State University as a graduate student. His Ph.D. thesis has focused on macroeconomic theory and political economy.