Essays in Public and Labor Economics

Inauguraldissertation zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Dipl.-Volksw. Max Löffler
aus
Halle an der Saale

Köln 2018

Erstgutachter: Prof. Dr. Felix Bierbrauer
Zweitgutachter: Prof. Dr. Andreas Peichl
Vorsitzender: Prof. Dr. Erik Hornung
Datum der Promotion: 13. Juli 2018

Acknowledgments

I am deeply grateful to my advisors, Felix Bierbrauer and Andreas Peichl, for introducing me to the science and art of economics. Thank you for providing an inspiring and challenging research environment and for your steady encouragement and mentoring. I greatly benefited from (and enjoyed) our discussions on economics, research, public policy, and politics.

This thesis is the result of five and a half years of research, most of which I spent working on joint projects with Sebastian Siegloch and Andreas Lichter. Our collaboration and our countless debates massively improved my understanding of economics and empirical research. The work with you has always been a pleasure to me, as have our procrastinating lunch debates. Thank you so much for sharing the hopes and doubts and the ups and downs during the past years.

I would like to thank my co-authors Philipp Dörrenberg and Denvil Duncan for introducing me to the field of experimental economics and their valuable feedback on other projects. I also benefited from many comments, questions, and suggestions of my colleagues Amelie Schiprowski, Carina Neisser, Emanuel Hansen, Eric Sommer, Erik Hornung, Johannes Pfeifer, Marius Vogel, Mathias Dolls, Nico Pestel, and Paul Hufe. Finally, I am grateful to IZA Bonn, ZEW Mannheim, and CMR in Cologne for their support during my doctoral studies.

Contents

List of Tables

List of Figures

Introduction

This thesis consists of four independent chapters. The first two chapters cover topics in the fields of public economics and political economy, respectively. Chapters three and four focus on questions from the field of labor economics.

In Chapter 1, jointly written with Sebastian Siegloch, we analyze the incidence and the welfare implications of property taxation. We suggest a novel theoretical perspective by introducing property taxes in a spatial equilibrium model, where workers and firms are mobile but have location-specific preferences, and where tax revenues finance local public goods. The model predicts that welfare effects of property taxation depend on four reduced-form elasticities. We estimate these elasticities using an event-study design and exploiting the institutional setting of municipal property taxation in Germany with more than 31,000 tax reforms in the years between 1992–2017. We simulate the welfare implications of tax increases and find that renters bear one fifth, firm owners around one third, and land owners more than 40 percent of the welfare loss. Our study adds to the existing literature on the effects of property taxation, which has offered a wide range of incidence estimates, ranging between 0-115 percent. The event-study results also highlight the dynamics of the property tax incidence, an important but so far neglected issue.

In Chapter 2, co-authored with Andreas Lichter and Sebastian Siegloch, we investigate the long-run effects of government surveillance on trust and economic performance. We study the case of the Stasi in socialist East Germany, which implemented one of the largest state surveillance systems of all time. Exploiting regional variation in the number of spies and the specific administrative structure of the system, we combine a border discontinuity design with an instrumental variables approach to estimate the long-term causal effect of government surveillance after the fall of the Iron Curtain. We find that a larger spying density in the population led to persistently lower levels of interpersonal and institutional trust in post-reunification Germany. We also find evidence of substantial and long-lasting economic effects of Stasi spying, resulting in lower income and higher exposure to unemployment. Our study contributes to the steadily growing literature on the relationship between institutions, culture, and economic performance. We confirm the long-term positive effects of institutional quality on economic performance, highlighting the importance of trust, social capital, and social ties for economic prosperity.

In Chapter 3, co-authored with Philipp Dörrenberg and Denvil Duncan, we test whether labor supply responds symmetrically to wage increases and decreases using a randomized field experiment with workers on Amazon's Mechanical Turk. The results show that wage increases have smaller effects on labor supply than wage decreases of equal magnitude, especially on the extensive margin where the elasticity for a wage decrease is twice that for a wage increase. This finding suggests that labor supply responses to non-marginal wage changes are asymmetric. As many studies in the labor supply literature exploit both positive and negative variation in wages to estimate an average wage elasticity of labor supply, our results suggest that existing estimates likely overstate the effect of wage increases and underestimate the effect of wage decreases. Our study further raises questions about the comparability of labor supply elasticities across studies that differ in the sign of the wage changes used for identification. We discuss the potential mechanisms behind our results including standard models of labor supply, loss aversion, and reciprocity.

In Chapter 4, jointly written with Andreas Peichl and Sebastian Siegloch, we systematically investigate the sensitivity of structural labor supply models with respect to underlying modeling choices. The analysis is motivated by the considerable dispute among researchers about the magnitude of labor supply elasticities. While differences in estimates – especially between micro and macro models – have been recently attributed to frictions and adjustment costs, we show that the variation in elasticities derived from structural labor supply models can also be explained by modeling assumptions. To this end, we estimate 3,456 different models on the same data each representing a plausible combination of frequently made choices. While many modeling assumptions do not systematically affect labor supply elasticities, our controlled meta-analysis shows that results are very sensitive to the treatment of hourly wages in the estimation. For example, different (sensible) choices concerning the modeling of the underlying wage distribution and especially the imputation of (missing) wages lead to estimates of the labor supply elasticity between 0.2 and 0.65. We hence conclude that researchers should pay more attention to the robustness of their estimations with respect to the wage treatment. Our findings have important policy implications as labor supply elasticities are key parameters when evaluating or designing optimal tax benefit policies.

Chapter 1

Property Taxation, Housing, and Local Labor Markets*

1.1 Introduction

Property taxes are an important instrument for governments around the world. Understanding how they affect local housing and labor markets is important to design optimal policies. In light of the recent surge in house prices especially in growing cities, knowledge of the incidence and the welfare implications of property taxation seems more crucial than ever. Despite more than a century of economic research,[1] our understanding of the effects of property taxes is still in a "sad state" (Oates and Fischel, 2016, p. 415) and we can only speculate about the incidence of property taxation. There are two main reasons for this.

First, competing theoretical models with quite different perspectives on local property taxation exist. On the one hand, the *capital tax view* adopts a general equilibrium perspective and argues that the national average burden of the property tax is borne by capital owners, i.e., landlords (Mieszkowski, 1972, Mieszkowski and Zodrow, 1989). Only local deviations from the national average are passed on to renters. On the other hand, the *benefit view* builds on a Tiebout (1956) model with perfect zoning and mobile individuals, who choose among municipalities offering different combinations of tax rates and local public goods (Hamilton, 1975, 1976). In this type of models, the tax is equivalent to a user fee for local public services, whereas the tax is progressive, falling mainly on richer landlords in the capital tax view.

Second, from an empirical point of view, identifying the impact of property taxes on local jurisdictions is challenging for various reasons. There is a general lack of high-quality data on property taxes, public services, and residential dwellings with a sufficiently large number of observations. Oftentimes only cross-sectional data is available, where taxes and services vary simultaneously. Another complication arises as municipalities may not only differ in their

* This chapter is based on unpublished joint work with Sebastian Siegloch.

[1] See, e.g., the recent surveys by Zodrow (2001), Oates and Fischel (2016) and England (2016) for an overview.

tax rates, but also in their assessment practice of property values, which renders the tax base endogenous (as, for example, in the US, see Palmon and Smith, 1998). Last, most papers adopt a partial equilibrium perspective and only look at quantity or price effects – one exception being a recent paper by Lutz (2015) who looks at both capitalization and capital investment.

In this paper, we readdress the question of who bears the burden of property taxes. Theoretically, we make a novel contribution by introducing property taxation into a local labor market model (Moretti, 2011). In this spatial equilibrium model, individuals and firms pay property taxes, are mobile and respond to local prices and amenities as assumed in Tiebout sorting models. At the same time, workers have location-specific preferences and firms vary in productivity across places, which limits the regional mobility of both agents. We introduce a construction sector, which produces residential and commercial floor space following Ahlfeldt et al. (2015). Importantly, we incorporate benefit view elements by allowing local governments to use tax revenues to finance local public goods. We derive theoretical predictions for the incidence of property taxes on tenants, land owners, and firm owners, and provide simple formulas for the marginal welfare effects of property tax changes, that are governed by a few estimable price elasticities.[2]

In the second part of the paper, we test the theoretical predictions using rich administrative panel data and the quasi-experimental setting of property taxation in Germany (*Grundsteuer*). Municipalities may autonomously adjust local property tax rates via scaling factors that they can set at their own discretion. Importantly, and in contrast to other tax systems, municipalities, however, cannot influence the assessment of property values, which is conducted by states tax offices. All legal rules determining the tax burden are set at the federal level and cannot be influenced by municipalities, either. Hence, municipalities can only set the tax rates, which is important for the identification of causal effects. We gathered administrative data for the universe of 8,481 West German municipalities between 1992–2017. Each year, more than ten percent of the municipalities change their local property tax, resulting in a maximum of 31,862 tax reforms that we can exploit for identification. As house price data is much harder to obtain, we rely on a smaller sample with house price and net rent indices for different construction types and qualities for 436 (mostly urban) municipalities. We restrict most of our analysis to this subsample covering roughly 40 percent of the population.

We implement a series of event studies exploiting within-municipality variation in tax rates over time to estimate the effects of property taxes on housing and land prices, the housing stock, population levels, land use, and wages. Our empirical model enables us to assess the dynamics

[2] We use the term incidence in the strict sense describing the effect of taxes on prices. We use the term pass-through as a synonym. We use our incidence estimates to simulate marginal welfare effects of the tax, which measure the share of the welfare loss borne by the respective agents.

of the treatment effect in the short and medium run (up to five periods after the tax reform). In addition, we can test the exogeneity of tax reforms with respect to the outcomes of interest by investigating pre-trends. In the absence of a pre-trend, the identifying assumption is that there is no systematic regional factor driving both municipal property tax rates and outcome variables. We explicitly test this assumption by flexibly controlling for shocks at the commuting zone level and find that estimates are robust.

Our spatial equilibrium model shows that the tax incidence depends on the relative size of the effective housing supply and housing demand elasticities, which mirrors the stylized textbook result of tax incidence in partial equilibrium models. The tax burden is fully shifted on renters if housing supply is sufficiently elastic. As suggested by the *benefit view* literature, the compensating, negative effects of property taxes on net rents are mitigated if property tax increases translate into higher public good provision. In addition, our model predicts that municipal population decreases in the medium run as cities become less attractive when property taxes increase. In a similar vein, capital investment, the housing stock, land use, and land prices decrease after tax increases. This corresponds to the *new view* result that property taxes distort the location of capital. Last, we show that local wages might increase following a tax increase, partly compensating workers for rising costs of living.

The empirical results confirm most of the theoretical priors. We show that real net rents decrease in the short run – implying that part of the tax burden is on the landlord – but start to revert back to the pre-reform level three years after a tax increase. This suggests that both the statutory and the economic incidence of the tax are on the tenant in the long run. As predicted by the model, both municipal population and the housing stock respond negatively to higher local property taxes, reflecting the fact that higher costs of living make a city less attractive. The same pattern holds for land sales and land prices. However, we do not find significant effects on local wages. We also show that house prices, land prices, wages, population levels, the housing stock, and land sales do not change systematically prior to a tax change, which suggests that reverse causality is not an issue.

Linking the empirical results to the theoretical model, we calculate marginal welfare effects of property tax increases borne by tenants, firm owners, and land owners, respectively. Our simulation results show that workers bear one fifth and firm owners one third of the welfare loss. The remaining half of the burden is borne by land owners, who thus face the largest loss in welfare. Importantly, the welfare implications hardly depend on whether or not and to which degree property tax increases are mirrored by rising local public good levels. These results also highlight the importance of going beyond the pure economic tax incidence and studying the welfare implications of property tax increases.

Related Literature. We add to the literature by analyzing the effects of property taxes in a local labor market framework, which has become the "workhorse of the urban growth literature" (Glaeser, 2009, p. 25). In recent years, the traditional Rosen-Roback model (Rosen, 1979, Roback, 1982) has been extended to account for location-specific preferences of workers and differential productivity of firms, which relax the perfect mobility assumption in traditional models (Moretti, 2011, Kline and Moretti, 2014, Suárez Serrato and Zidar, 2016). We further add to the literature by endogenizing the supply of developed land, and incorporating a construction sector as in Ahlfeldt et al. (2015). Moreover, we introduce endogenous amenities by allowing local governments to spend the property tax revenue on local public goods.

Our framework allows for capitalization into local prices while workers' utility might still differ across places in equilibrium, other than in Brueckner (1981). While our model is close to a capital tax world with endogenous amenities, we deviate from the assumption of a fixed capital stock in the economy. In contrast, we assume global capital markets and perfect mobility of capital. As a consequence, higher property taxes reduce the overall capital stock in the society, a channel that has been neglected in the previous literature (Oates and Fischel, 2016). Our model further implies a second type of capital, namely floor space, which is consumed by workers and used as input in firms' production. The housing stock is provided by a perfectly competitive construction sector (see Thorsnes, 1997, Epple et al., 2010, Combes et al., 2016). As in classical property tax studies, our model predicts that land owners will bear a substantial share of the property tax burden via lower prices and reduced demand for developed land.

Empirically, we provide evidence on the effects of the property tax on housing and labor market prices and quantities using administrative data from German municipalities. In particular, we add to the existing empirical literature on the property tax incidence on rents, which has predominantly focused on the US. Using Germany as a case study is particularly interesting in this context, as it has one of the highest renter rates and one of the largest private rental markets among Western countries. The previous literature has offered a wide range of estimates of the property tax incidence on rents: Orr (1968, 1970, 1972), Heinberg and Oates (1970), Hyman and Pasour (1973), Dusansky et al. (1981), and Carroll and Yinger (1994) estimate that between 0-115 percent of the tax burden is shifted onto renters. Our results show the dynamics of the property tax incidence in the short and medium run, an important issue given time lags in housing market adjustments (England, 2016). We find that property tax increases lead to lower house prices, which is evidence of capitalization into house values (Palmon and Smith, 1998, de Bartolomé and Rosenthal, 1999). Our findings of a negative effect on municipal population levels are in line with evidence provided by Ferreira (2010) and Shan (2010), who show that property taxes affect mobility rates of the elderly. Last, our study offers evidence that property tax increases reduce housing investment, a mechanism identified by Lyytikäinen (2009) for the case of Finnish municipalities. In a similar vein, Lutz (2015) finds that property taxes reduce building permits and capital investment in rural areas, while they are capitalized into land prices in urban areas.

The remainder of this paper is organized as follows. In Section 1.2 we set up our theoretical model. Section 1.3 presents the institutional framework of property taxation in Germany. Section 1.4 provides information on the used data. We set up our empirical model in Section 1.5. In Section 1.6, we present our reduced-form results. Section 1.7 discusses welfare effects of the tax, Section 1.8 concludes.

1.2 Theoretical Model

We introduce local property taxation in a Rosen-Roback type general equilibrium model of local labor markets (Moretti, 2011, Kline and Moretti, 2014, Suárez Serrato and Zidar, 2016). The model consists of four groups of agents: workers, firms producing tradable goods, construction companies producing floor space, and land owners. Workers and firms are mobile and locate in one out of C cities, indexed by c.

First, we outline the model in Sections 1.2.1–1.2.5. Second, we solve for the spatial equilibrium and use comparative statics to show how changes in the property tax rate affect the equilibrium outcomes, i.e., population size, floor space, land use, rents, wages and land prices (see Section 1.2.6). In Section 1.2.7, we derive the welfare effects of tax changes and show how marginal welfare effects relate to the key elasticities of the model. Appendix 1.B provides a more comprehensive description of the model including all derivations.

1.2.1 Workers

There is a continuum of $N = 1$ workers indexed by i. Labor is homogeneous and each worker provides inelastically one unit of labor, earns a wage w_c, and pays rent r_c^H for residential floor space. In the theoretical model, we assume that there is only one homogeneous housing good and do not differentiate between owner-occupied and rental housing. In equilibrium, the implicit price of rental and owner occupied housing has to be the same (Poterba, 1984).[3] Workers maximize utility over floor space consumption h_i, a composite good bundle x_i of non-housing goods, whose aggregate price is normalized to one, and locations c. Workers are mobile across municipalities, but individuals have idiosyncratic location preferences e_{ic}, such that local labor supply is not necessarily infinitely elastic. In addition, there is an exogenous city-specific consumption amenity A_c, related to climate and geography, for instance. Each

[3] Empirically, we will look separately at the effects of rents and house sales price.

city c levies a municipal property tax denoted by t_c, with the statutory incidence on the user of the housing service.[4] Local governments spend (part of) the tax revenue on (improvements) of local public goods G_c.

Workers maximize utility $U_{ic} = A_c G_c^{\delta} \left(h_i^{\alpha} x_i^{1-\alpha}\right)^{1-\delta} e_{ic}$, subject to the budget constraint $r_c^H(1+t_c)h_i + px_i = w_c$, which yields indirect utility:

$$V_{ic}^H = a_0 + \underbrace{(1-\delta)\left(\ln w_c - \alpha \ln r_c^H - \alpha \ln[1+t_c]\right) + \delta \ln G_c + \ln A_c}_{=V_c^H} + \ln e_{ic}$$

where α denotes the housing share in consumption, individuals have preferences $\delta \in (0,1)$ for the public vs. private goods, and a_0 is a constant. Indirect utility can be rewritten as the sum of a city-specific part V_c^H and a worker-location-specific component e_{ic}. In line with the literature on local labor markets, we assume that the logarithm of e_{ic} is independent and identically extreme value type I distributed with scale parameter $\sigma^H > 0$. Hence, the greater σ^H, the stronger workers' preferences for given locations and the lower their mobility.

Given that the total number of workers is normalized to one and the number of cities C is large, log labor supply in municipality c is given by:

$$\ln N_c^S = \underbrace{\frac{1-\delta}{\sigma^H}}_{=\epsilon^{\text{NS}}} \ln w_c - \underbrace{\frac{\alpha(1-\delta)}{\sigma^H}}_{=1+\epsilon^{\text{HD}}} \ln r_c^H - \underbrace{\frac{\alpha(1-\delta)}{\sigma^H}}_{=1+\epsilon^{\text{HD}}} \ln \tau_c + \underbrace{\frac{\delta}{\sigma^H}}_{=\delta\epsilon^{\text{A}}} \ln G_c + \underbrace{\frac{1}{\sigma^H}}_{=\epsilon^{\text{A}}} \ln A_c + a_1 \tag{1.1}$$

where a_1 is a constant and we redefine the property tax rate as $\tau_c = 1 + t_c$. Equation (1.1) also defines various key elasticities of our model, such as the labor supply elasticity $\epsilon^{\text{NS}} = \frac{1-\delta}{\sigma^H}$.

Demand for residential housing in city c is determined by the number of workers in city c and their individual housing demand:

$$\ln H_c = \ln N_c + \ln \alpha + \ln w_c - \ln r_c^H - \ln \tau_c. \tag{1.2}$$

It follows that the intensive margin housing demand elasticity conditional on location choice is equal to -1. In addition, there is an extensive margin with people leaving the city in response to higher costs of living. The aggregate residential housing demand elasticity is given by:

$$\frac{\partial \ln H_c}{\partial \ln r_c^H} = \frac{\partial \ln N_c}{\partial \ln r_c^H} - 1 = -\frac{\alpha(1-\delta) + \sigma^H}{\sigma^H} = \epsilon^{\text{HD}} < 0.$$

[4] For simplicity, we assume that property is taxed *ad valorem*. Our main theoretical prediction regarding the tax incidence is however unchanged when modeling the property tax as a specific tax instead (see Appendix 1.B.9).

1.2.2 Firms

There are F firms indexed by j that operate under monopolistic competition and produce a tradable output good Y_{jc}, using labor N_{jc} and commercial floor space M_{jc}. Similar to the worker problem there is an exogenous local production amenity B_c and idiosyncratic firm-location-specific productivity shifters ω_{jc}. The price of the output good is denoted by p_{jc}. Following Melitz (2003), we can use total product demand Q to write firms profits as:

$$\Pi_{jc}^F = Q^{1-\rho} \underbrace{\left(B_c \omega_{jc} N_{jc}^{\beta} M_{jc}^{1-\beta} \right)^{\rho}}_{=Y_{jc}} - w_c N_{jc} - r_c^M \tau_c \kappa M_{jc}$$

where β denotes the labor share in production, r_c^M is the factor price of commercial floor space, $\kappa > 0$ is a scale parameter that allows property taxes on commercial rents to differ from residential property taxes, and the constant product demand elasticity is given by $\epsilon^{\mathrm{PD}} = -\frac{1}{1-\rho}$.

Profit maximization yields optimal factor demands, N_{jc}^{D*}, M_{jc}^{D*}, conditional on local productivity and factor prices. Using the profit function from above, the value of firm j in city c in terms of factor costs and local productivity is then given by (Suárez Serrato and Zidar, 2016):

$$V_{jc}^F = \frac{1-\rho}{\rho} \ln \Pi_{jc}^F \left(N_{jc}^{D*}, M_{jc}^{D*} \right) = b_0 + \underbrace{\ln B_c - \beta \ln w_c - (1-\beta) \ln r_c^M - (1-\beta) \ln(\tau_c \kappa)}_{=V_c^F} + \ln \omega_{jc}$$

where b_0 is a constant. Assuming that idiosyncratic productivity shifters $\ln \omega_{jc}$ are drawn i.i.d. from an extreme value type I distribution with scale parameter σ^F, we can derive the number of firms in city c, normalizing the total number of firms to $F = 1$. Using the number of firms in a given city, which defines the extensive margin of labor demand, and optimal labor demand conditional on location at the intensive margin (N_{jc}^{D*}), aggregate labor demand in city c is given by:

$$\begin{aligned}\ln N_c^D = & \underbrace{\left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right)}_{=\epsilon^{\mathrm{B}}} \ln B_c \underbrace{- \left(1 + \beta \left[\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right] \right)}_{=\epsilon^{\mathrm{ND}}} \ln w_c \underbrace{-(1-\beta) \left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right)}_{=1+\epsilon^{\mathrm{MD}}} \ln r_c^M \\ & \underbrace{-(1-\beta) \left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right)}_{=1+\epsilon^{\mathrm{MD}}} \ln(\tau_c \kappa) + b_1 \end{aligned} \tag{1.3}$$

where b_1 is a constant. The labor demand elasticity is defined as:

$$\frac{\partial \ln N_c^D}{\partial \ln w_c} = \underbrace{-\frac{\beta}{\sigma^F}}_{\text{Ext. margin}} \underbrace{-1-\frac{\beta\rho}{1-\rho}}_{\text{Int. margin}} = \epsilon^{\text{ND}} < 0.$$

Analogously, we can derive the demand for commercial floor space using the intensive margin demand, M_{jc}^{D*}, and the location choice of firms:

$$\begin{aligned}\ln M_c^D = & \underbrace{\left(\frac{1}{\sigma^F}+\frac{\rho}{1-\rho}\right)}_{=\epsilon^{\text{B}}}\ln B_c \underbrace{-\beta\left(\frac{1}{\sigma^F}+\frac{\rho}{1-\rho}\right)}_{=1+\epsilon^{\text{ND}}}\ln w_c \underbrace{-\left(1+[1-\beta]\left[\frac{1}{\sigma^F}+\frac{\rho}{1-\rho}\right]\right)}_{=\epsilon^{\text{MD}}}\ln r_c^M \\ & \underbrace{-\left(1+[1-\beta]\left[\frac{1}{\sigma^F}+\frac{\rho}{1-\rho}\right]\right)}_{=\epsilon^{\text{MD}}}\ln(\tau_c\kappa)+b_2 \end{aligned} \tag{1.4}$$

with constant b_2. The commercial floor space demand elasticity is defined as:

$$\frac{\partial \ln M_c^D}{\partial \ln r_c^M} = -\frac{1-\beta}{\sigma^F} - 1 - \frac{\rho(1-\beta)}{1-\rho} = \epsilon^{\text{MD}} < 0.$$

1.2.3 Construction Sector

We assume that a competitive, local construction sector provides both residential and commercial floor space. For positive supply on the two markets, there must be a no-arbitrage condition between both construction types. Following Ahlfeldt et al. (2015), we assume that the residential share μ of total floor space is determined by the price of residential housing, r_c^H, and commercial floor space, r_c^M:

$$\begin{aligned} &\mu = 1, && \text{for } r_c^M < \phi r_c^H \\ &\mu \in (0,1), && \text{for } r_c^M = \phi r_c^H \\ &\mu = 0, && \text{for } r_c^M > \phi r_c^H \end{aligned} \tag{1.5}$$

with $\phi \geq 1$ denoting additional regulatory costs of commercial land use compared to residential housing.[5] In equilibrium, the no-arbitrage condition fixes the ratio between residential and commercial floor space prices and every municipality has positive supply of residential housing H_c and commercial floor space M_c. We can rewrite the two types of floor space in terms of

[5] We abstract from heterogeneity in the residential land use share, μ, and the regulatory markup, ϕ, for simplicity. This assumption does not influence our results qualitative.

total floor space, S_c, available in city c:

$$H_c = \mu S_c \qquad\qquad M_c = (1-\mu)S_c. \tag{1.6}$$

We follow the standard approach in urban economics and assume that the housing construction sector relies on a Cobb-Douglas technology with constant returns to scale using land ready for construction L_c and capital K_c to produce total floor space (see, e.g., Thorsnes, 1997, Epple et al., 2010, Combes et al., 2016):

$$S_c = H_c + M_c = L_c^{\gamma} K_c^{1-\gamma} \tag{1.7}$$

with γ being the output elasticity of land. We assume global capital markets with unlimited supply at an exogenous rate s. Profits in the construction industry are given by $\Pi_c^C = r_c^M S_c - l_c L_c - sK_c$, where l_c denotes the price of land. Capital demand is then given by:

$$\ln K_c = \frac{1}{\gamma}\ln(1-\gamma) + \frac{1}{\gamma}\ln r_c^M + \ln L_c - \frac{1}{\gamma}\ln s \tag{1.8}$$

which can be used to solve for the price ratio of land to floor space in city c:

$$\ln l_c = c_0 - \frac{1-\gamma}{\gamma}\ln s + \frac{1}{\gamma}\ln r_c^M \tag{1.9}$$

with c_0 being a constant. Land prices increase in the commercial floor space rent r_c^M (and equivalently in residential rents r_c^H).

1.2.4 Land Supply

While the total land area in each municipality is fixed and inelastic, the share of land ready for residential or commercial construction may be elastic. We model land supply in city c as:

$$\ln L_c = \theta \ln l_c \tag{1.10}$$

with the supply elasticity of land ready for building defined as $\epsilon^{\text{LS}} = \theta > 0$. We model the land supply elasticity as constant across places for simplicity. In the empirical part of the paper we test for heterogeneous effects by geographical supply determinants. In line with the literature, we assume that landowners are absent (see, e.g., Kline and Moretti, 2014, Ahlfeldt et al., 2015, Diamond, 2016).

1.2.5 Local Governments

Local governments use share $\psi \in (0,1)$ of the property tax revenue to finance the local public good G_c. All remaining revenues are distributed lump-sum to all workers in the economy irrespective of location (share $1-\psi$). The government budget is defined as:

$$G_c = \psi \underbrace{\left(H_c r_c^H t_c + M_c r_c^M \left[\{1+t_c\}\kappa - 1\right]\right)}_{\text{Total tax revenue}}, \tag{1.11}$$

where total tax revenue is the sum of residential property taxes, $H_c r_c^H t_c$, and property taxes on rented commercial floor space, $M_c r_c^M$. Increases in city c's property tax rate t_c yield higher tax revenues, leading to a mechanical increase in local spending on the public good.

1.2.6 Equilibrium and Comparative Statics

The spatial equilibrium is described by equations (1.1) through (1.11). It is determined by equalizing supply and demand on the markets for labor, residential housing, commercial floor space, and land in each city, accounting for the government budget constraint. The solution to this system of equations yields equilibrium quantities in terms of population, residential housing, commercial floor space, use of capital, and developed land as well as equilibrium prices for labor, residential housing, commercial floor space, and land, which are derived in Appendix 1.B.6.

In the following, we show how equilibrium outcomes respond to changes in property taxes. The comparative statics of our model yield theoretical predictions on the impact of property taxes on equilibrium quantities and prices that eventually govern the welfare effects (see Section 1.2.7). We estimate and test quantity and price responses against the theoretical priors using the institutional setting in Germany in Section 1.6. Proposition 1 summarizes the price effects of property tax increases in our model.

Proposition 1 (Price Effects)**.** *Let r_c^{H*}, r_c^{M*}, l_c^*, and w_c^*, respectively, denote the equilibrium net rent for residential housing, the net rent for commercial floor space, the land price, and the wage level in municipality c, each as a function of the local property tax rate τ_c and equilibrium public good provision $G_c^*(\tau_c)$. An increase in city c's tax rate triggers two effects on equilibrium prices:*

(i) A direct, negative effect on residential and commercial rents as well as the land price, and a direct effect on the local wage level that is theoretically ambiguous and may be positive or negative.

(ii) An indirect effect operating through the capitalization of public goods, which moderates

the negative effect on housing and land prices as long as tax increases raise the public good supply.

$$\frac{d\ln r_c^{H*}\left(\tau_c, G_c^*[\tau_c]\right)}{d\ln\tau_c} = \underbrace{\frac{\partial\ln r_c^{H*}}{\partial\ln\tau_c}}_{<0} + \underbrace{\frac{\partial\ln r_c^{H*}}{\partial\ln G_c^*}}_{>0}\frac{\partial\ln G_c^*}{\partial\ln\tau_c} = \frac{d\ln r_c^{M*}\left(\tau_c, G_c^*[\tau_c]\right)}{d\ln\tau_c} \tag{1.12a}$$

$$\frac{d\ln l_c^*\left(\tau_c, G_c^*[\tau_c]\right)}{d\ln\tau_c} = \underbrace{\frac{\partial\ln l_c^*}{\partial\ln\tau_c}}_{<0} + \underbrace{\frac{\partial\ln l_c^*}{\partial\ln G_c^*}}_{>0}\frac{\partial\ln G_c^*}{\partial\ln\tau_c} \tag{1.12b}$$

$$\frac{d\ln w_c^*\left(\tau_c, G_c^*[\tau_c]\right)}{d\ln\tau_c} = \underbrace{\frac{\partial\ln w_c^*}{\partial\ln\tau_c}}_{\lesseqgtr 0} + \underbrace{\frac{\partial\ln w_c^*}{\partial\ln G_c^*}}_{<0}\frac{\partial\ln G_c^*}{\partial\ln\tau_c}. \tag{1.12c}$$

Proof. See Section 1.B.7 in the Appendix. □

The direct effect decreases equilibrium rental prices for residential and commercial floor space and thereby compensates for rising taxes. The equilibrium price of land, being the inelastic input factor in the floor space production, also decreases in response to tax increases while holding public goods constant. As shown in the Appendix, the model intuitively fixes the ratio of these direct marginal effects to be equal to the land share in the floor space production, $\frac{\partial\ln r_c^{H*}}{\partial\ln\tau_c}\Big/\frac{\partial\ln l_c^*}{\partial\ln\tau_c} = \gamma$. The direct effect on rents closely mirrors the tax incidence in a standard partial equilibrium model.

Corollary 1. *The direct rent response to property tax increases in the spatial equilibrium is determined by the effective housing supply and demand elasticities, which also account for equilibrium responses on the land and the labor market as well as the market for commercial floor space:*

$$\frac{\partial\ln r_c^{H*}}{\partial\ln\tau_c} = \frac{\tilde{\epsilon}^{\mathrm{HD}}}{\tilde{\epsilon}^{\mathrm{HS}} - \tilde{\epsilon}^{\mathrm{HD}}} < 0 \tag{1.13}$$

where $\tilde{\epsilon}^{\mathrm{HS}}$ *and* $\tilde{\epsilon}^{\mathrm{HD}}$ *denote the effective housing supply and demand elasticities, respectively.*

Proof. See Lemma 1.B.2 in Appendix 1.B.7. □

The more elastic the effective housing supply, the lower the tax burden for the supply side, i.e., landlords and constructors. The more elastic effective housing demand, the larger the compensating effects on net rents and wages, and the lower the tax burden on renters.

The direct effect on wages is ambiguous. On the one hand, higher property tax payments raise the factor price of commercial floor space and reduce firms' floor space demand, which decreases the marginal product of labor. On the other hand, property taxes increase worker's costs of living, which – given worker mobility – induces a demand for higher wages as a compensating differential. Hence, the sign of the direct wage effect is determined by the relative sizes of the commercial and residential floor space demand. The direct effect of property taxes on the local real wage, defined as the wage over costs of living in city c, is unambiguously negative as shown in Lemma 1.B.4 in the Appendix.

The total effects on equilibrium prices differ from the direct effects because tax increases raise additional revenues that the local government (partly) spends for additional supply of local public goods. Public goods are one of the determinants of workers' location choice and thus affect labor supply and residential housing demand. If higher taxes increase the local level of public goods, which is the case if initial tax rates are not too high, there is an indirect effect that alleviates the direct, compensating effect on rents, land price, and wages (if the direct wage effect is positive). The magnitudes and signs of the total effects depend on the relative importance of direct and indirect effects.

Quantity effects of property tax increases follow immediately from the model outlined above.

Lemma 1 (Quantity Effects). *Let H_c^* and M_c^* denote the residential housing stock and the commercial floor space in equilibrium, respectively, let L_c^* be the equilibrium land area used for development and let N_c^* be the equilibrium population level in municipality c, each as a function of the property tax rate τ_c and equilibrium public good provision $G_c^*(\tau_c)$. An increase in city c's property tax rate t_c triggers (i) a direct, negative, and (ii) an indirect, potentially moderating effect on equilibrium quantities:*

$$\frac{d\ln H_c^*\left(\tau_c, G_c^*[\tau_c]\right)}{d\ln\tau_c} = \underbrace{\frac{\partial\ln H_c^*}{\partial\ln\tau_c}}_{<0} + \underbrace{\frac{\partial\ln H_c^*}{\partial\ln G_c^*}}_{>0}\frac{\partial\ln G_c^*}{\partial\ln\tau_c} = \frac{d\ln M_c^*\left(\tau_c, G_c^*[\tau_c]\right)}{d\ln\tau_c} \tag{1.14a}$$

$$\frac{d\ln L_c^*\left(\tau_c, G_c^*[\tau_c]\right)}{d\ln\tau_c} = \underbrace{\frac{\partial\ln L_c^*}{\partial\ln\tau_c}}_{<0} + \underbrace{\frac{\partial\ln L_c^*}{\partial\ln G_c^*}}_{>0}\frac{\partial\ln G_c^*}{\partial\ln\tau_c} \tag{1.14b}$$

$$\frac{d\ln N_c^*\left(\tau_c, G_c^*[\tau_c]\right)}{d\ln\tau_c} = \underbrace{\frac{\partial\ln N_c^*}{\partial\ln\tau_c}}_{<0} + \underbrace{\frac{\partial\ln N_c^*}{\partial\ln G_c^*}}_{>0}\frac{\partial\ln G_c^*}{\partial\ln\tau_c}. \tag{1.14c}$$

Proof. See Section 1.B.7 in the Appendix. □

Again, looking first at the direct effects, the model predicts the quantities on the floor space, land, and labor market to decrease in response to an increase in the property tax given the real-wage loss in city c. Workers thus leave the city, employment declines, construction and land use decrease.

The magnitude and the sign of the total effects of tax increases on equilibrium quantities again depend on the relative importance of the direct vs. the indirect effect, that operates through increases in local public good supply and moderates the direct effects as long as property tax increases lead to higher spending on local public goods.

1.2.7 Welfare Effects

In this section, we derive the marginal welfare effects of property tax changes for the different agents in the model. We present the results for representative agents in city c, implicitly assuming that the distribution of agents across municipalities is homogeneous (Suárez Serrato and Zidar, 2016). Proposition 2 summarizes the welfare effects based on price responses.

Proposition 2 (Welfare Effects)**.** *Let W^H, W^F, W^C, and W^L denote the welfare of workers, firm owners, constructors, and land owners in the spatial equilibrium, respectively. A marginal increase in city c's property tax rate t_c leads to welfare changes that are determined by:*

(i) the elasticities of equilibrium rents, land prices, and wages with respect to the property tax rate,

(ii) the responsiveness of the local public good provision in equilibrium with respect to the tax,

(iii) three exogenous model parameters, namely the housing share in consumption, α, the labor share in the tradable good production, β, and the preferences for local public goods, δ

$$\frac{dW^H}{d\ln\tau_c} = -\left([1-\delta]\left[\alpha + \alpha\frac{d\ln r_c^{H*}}{d\ln\tau_c} - \frac{d\ln w_c^*}{d\ln\tau_c}\right] - \delta\frac{d\ln G_c^*}{d\ln\tau_c}\right) \tag{1.15a}$$

$$\frac{dW^F}{d\ln\tau_c} = -\left([1-\beta] + [1-\beta]\frac{d\ln r_c^{M*}}{d\ln\tau_c} + \beta\frac{d\ln w_c^*}{d\ln\tau_c}\right) \tag{1.15b}$$

$$\frac{dW^C}{d\ln\tau_c} = 0 \tag{1.15c}$$

$$\frac{dW^L}{d\ln\tau_c} = \frac{d\ln l_c^*}{d\ln\tau_c}. \tag{1.15d}$$

Proof. See Section 1.B.8 in the Appendix. □

The analysis shows that workers' marginal welfare loss from tax hikes decreases in the preference for the local public good, δ. Hence, the stronger the preferences for public goods and the stronger the transmission of taxes into public good spending, the smaller the welfare loss as workers are compensated for rising costs of living.

Proposition 2 implies that the rent, land price, wage, and public good elasticities with respect to the property tax are sufficient to infer the welfare effects of the tax in a local labor market model – given the housing share in consumption, the labor share in production, and the preferences for public goods, which can be calibrated according to official statistics. In the following sections, we estimate the behavioral responses to changes in property taxes using the German institutional framework. In Section 1.7, we use these behavioral elasticities and the respective welfare formulas to assess the marginal welfare effects of the property tax.

1.3 Institutional Background

We test the theoretical predictions of our model using administrative data from German municipalities. In this section, we provide a short overview on the institutional setting of property taxation in Germany (see Spahn, 2004, for more details).

Property taxes are one of the oldest forms of taxation that is still used today. The current German property tax regulations are based on a law from 1936.[6] Besides local business taxes and municipal shares on federal income and sales taxes, the property tax is one of the three most important income sources for German municipalities. Property taxes account for around 15 percent of municipal revenues, amounting to a total of 12 billion EUR for all municipalities in 2013. All legal regulations of the German property tax, i.e., the definition of the tax base, as well as legal norms regarding the property assessment, are set at the federal level and have rarely been changed over the past decades.

The property tax liability is calculated according to the following formula, that we discuss in more detail below:

$$Tax\ Liability = Assessed\ Value \times \underbrace{Federal\ Tax\ Rate \times Municipal\ Scaling\ Factor}_{\text{Local Property Tax Rate}}. \tag{1.16}$$

Assessed Values. The house value (*Einheitswert*) is assessed by the tax offices of the federal state (not by the municipality) when the property is built and, importantly, remains fixed over time. There is no regular reassessment of properties to adjust the assessed value to the market value of the property or to inflation rates. Even when being sold, the assessed value does not change. Reassessments only occur if the owner creates a new building or substantially improves an existing structure on her land.[7] The last general assessment of property values in West

[6] The law distinguishes between taxes on agricultural land (*Grundsteuer A*) and taxes on residential and commercial land as well as improvements (*Grundsteuer B*). We focus solely on the latter one in this paper as only this type of the tax is relevant for residential housing markets and commercial floor space.

[7] The improvement has to concern the "hardware" of the property, such as adding a floor to the building. Maintaining the roof or installing a new kitchen does not yield reassessments. Lock-in effects or assessment limits are thus

Germany took place in 1964. In order to make the assessment comparable for buildings that have been constructed after that date, property values of new structures are evaluated on market rents as of 1964 using historical rent indices. So even new buildings are assessed as if they had been built several decades ago. As a consequence, assessed values differ substantially from current market values. Assessment notices do not provide any detail on how specific parts of the building contribute to the assessed value. This practice makes the assessment barely transparent for house owners, landowners and renters. The average assessed value for West German homes was 48,900 EUR in 2013, roughly a fifth of the reported current market value (according to the German Income and Expenditure Survey, EVS).

Federal Tax Rates. The federal tax rate (*Grundsteuermesszahl*) is set at 0.35 percent for all property types in West Germany with two exceptions (see Figure 1.1). First, single-family homes are taxed at 0.26 percent up to the value of 38,347 EUR; and at 0.35 percent for every Euro the house value exceeds this threshold. Second, two-family houses are taxed at 0.31 percent. All other property types are taxed at 0.35 percent. The federal tax schedule is thus progressive for single-family houses and otherwise flat. Once the property type has been determined by the state tax offices, land and structures are taxed at the same rate.[8] The average federal property tax rate in our sample is 0.32 percent.

Municipal Scaling Factors. Municipal councils decide yearly on the local scaling factor (*Hebesatz*). The decision is usually made in the last months of the preceding year, and most tax changes become effective on January 1st.

For a given housing stock and a fixed federal rate, local property tax rates only vary due to changes in local scaling factors. Figure 1.2 demonstrates the substantial cross-sectional and time variation in tax rates induced by changes in scaling factors. The left panel of the figure shows the local tax rates for all West German municipalities in 2017, assuming a federal tax rate of 0.32 percent. Local property tax rates vary between 0.73 and 1.71 percent (bottom and top one percent). Annual mean and median tax rates increased steadily from around 0.86 in 1992 to 1.17 percent in 2017. The average tax per square meter was 0.20 EUR for rental apartments, which corresponds to 3.29 percent on top of the average net rent in our sample.[9]

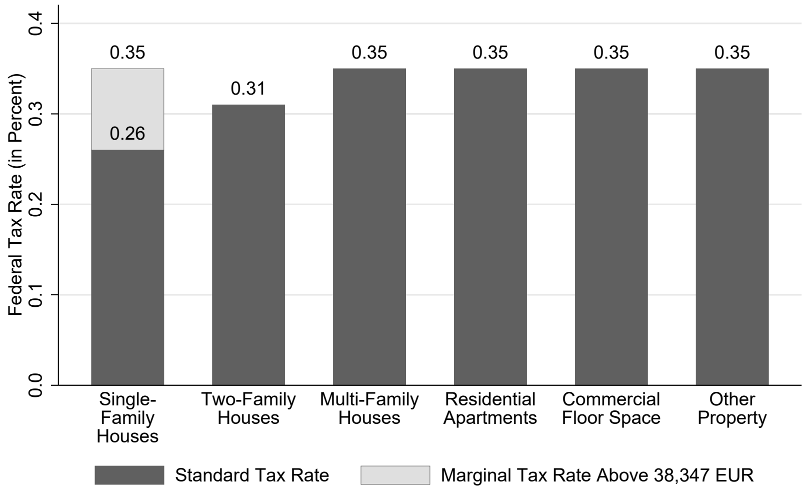
The right panel of Figure 1.2 demonstrates the number of municipal scaling factor changes in the period from 1992 to 2017. Over this period, more than ninety percent of all municipalities

not an issue in the German context other than in some US states (see, e.g., Ferreira, 2010, Bradley, 2017).

[8] As tax rates for developed properties and undeveloped land are also similar, the German system is thus essentially a one-rate property tax (see the discussion in Plassmann and Tideman, 2000, and Lyytikäinen, 2009, for the differences between one-rate and multi-rate tax systems and their effect on housing construction).

[9] The average tax burden for rental apartments in West Germany is published annually by the German Tenants' Association (*Deutscher Mieterbund*) based on a survey on operating costs (*Betriebskostenspiegel*).

Figure 1.1: Federal Tax Rates in West Germany (in Percent)



Notes: This graph shows the federal tax rates for different property types in West Germany. Federal tax rates are flat except for single-family houses, which are taxed at 0.26 percent up to an assessed value of 38,347 EUR and with a marginal tax rate of 0.35 percent above that threshold. All tax rates are in percent. The average federal tax rate in our sample is 0.32 percent. *Source:* § 15 *Grundsteuergesetz.*

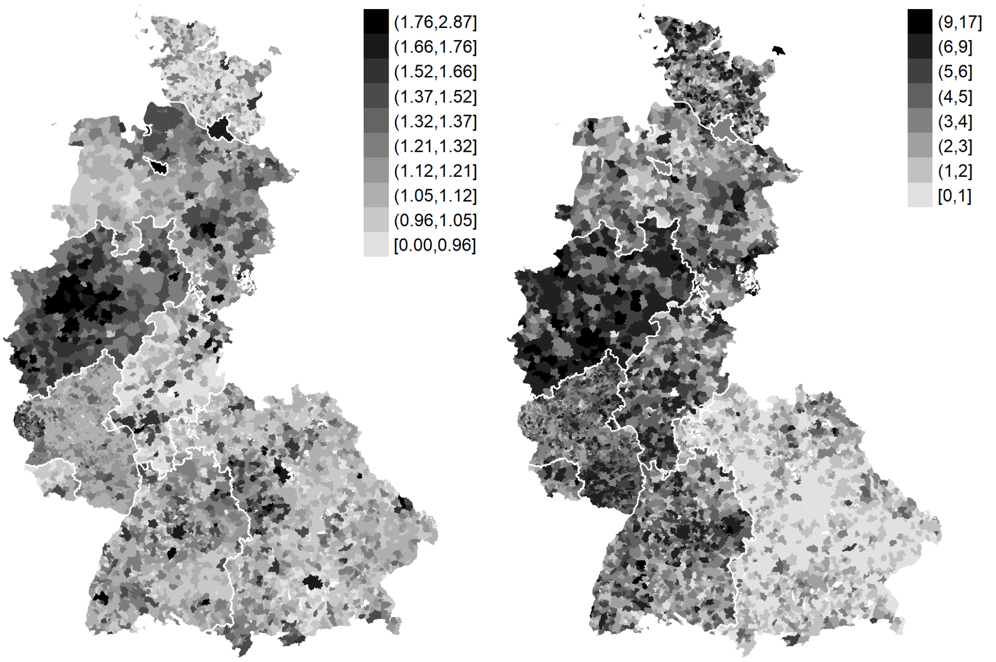
changed their local tax rate at least once, while less than six percent of municipalities still have the same multiplier as in the beginning of the 1990s. On average, municipalities changed the factor four times during this period, i.e., every six years. Many municipalities experienced even more changes. One percent of municipalities changed their property tax multiplier more than ten times since 1992. Almost 97 percent of all tax changes during this period are tax increases.

Statutory Incidence. Property owners are liable for the tax payment irrespective of whether the property is owner-occupied, for rent or vacant. However, for rental housing, property taxes are part of the ancillary costs that renters have to pay on top of net rents to their landlords according to the legal regulations on operating costs (*Betriebskostenverordnung*). In this regulation, landlords are directed to include the tax payments in the ancillary bill the renters receives each year and it is a common practice to do so. As a consequence, the statutory incidence is on the user of the housing service for both owner-occupied and rental housing.

Commercial Property Taxes. For German firms property tax liabilities are of second order. Municipalities' tax revenues from local business taxes were 43 billion EUR as of 2013, tax revenues from property taxes amounted to 12 billion EUR. From these 12 billion EUR, the largest share came from residential buildings. A conservative estimate is that two thirds of a

Figure 1.2: Variation in Local Property Tax Rates in West Germany

(a) Local Property Tax Rates in 2013 (in Percent) (b) Number of Property Tax Changes 1992–2017



Notes: The left panel of this figure shows the local property tax rates in 2013 for all West German municipalities, assuming a federal tax rate of 0.32 percent. The right panel depicts the number of local property tax changes by municipality in the period 1992–2017. Jurisdictional boundaries are as of December 31, 2010. White lines indicate federal state borders. *Source:* Federal Statistical Office and Statistical Offices of the federal states. *Maps:* © GeoBasis-DE / BKG 2015.

municipality's total area are for residential, one third for commercial use. Commercial property taxes thus make up less than ten percent of firms' total tax bill on average.

1.4 Data and Descriptive Statistics

This section gives an overview on the data used for our empirical analysis. We combine administrative data on the fiscal and economic situation of German municipalities with housing market information, and administrative wage data from social security registers (Section 1.4.1). In Section 1.4.2 we define our baseline data set used in the empirical analysis. Appendix 1.A provides more details on the definition and the sources of all variables.

1.4.1 Variables and Data Sources

German Municipality Data. We collect a comprehensive data set for all West German municipalities from the Federal Statistical Office and the Statistical Offices of the Länder, including data on the economic, fiscal and budgetary situation, most importantly local scaling factors, as well as population figures and information on the housing stock, land prices, land use, and local GDP. In addition we collect unemployment data from the Federal Employment Agency. The Federal Institute for Research on Building, Urban Affairs and Spatial Development provides us with definitions of commuting zones that are defined according to commuting flows (*Arbeitsmarktregionen*). Using these sources we construct a balanced panel for the universe of all 8,481 West German municipalities ranging from 1992 to 2017.[10] Table 1.A.1 in the Appendix provides details on the definition and data sources of all variables as well as the years for which data is available, Appendix Table 1.A.2 shows descriptive statistics.

Housing Price Data. We combine the municipality panel with housing market data provided by the German real estate association IVD (*Immobilienverband Deutschland*). This data set delivers eight distinct rent indices for standardized rental apartments with 70 square meter and three bedrooms, and seven house price indices for single-family buildings. These indices differ by construction year and apartment quality and thus allow us to study heterogeneous effects of property taxes. It is important to note, that this data only includes quoted net prices and quoted net rents (*Nettokaltmiete*) and does not contain information on operating costs, taxes, or actual transaction prices. Thus, we do not observe gross prices including property taxes.

We validate the house price and rent data against several other data sources: (i) official household survey data from the German micro census, which includes information on rents at the county level every four years, covering 89 large municipalities; (ii) housing market indicators provided by *Empirica*, an independent economic consultancy specialized in the real-estate sector, covering the same large municipalities over the period 2004–2013; (iii) data provided by *Bulwiengesa*, another real-estate consultancy, whose data set includes 102 municipalities. Figure 1.A.1 in the Appendix compares the different measures and shows that data quality of our IVD data set seems reasonable. In addition, Appendix Figure 1.A.2 plots average reported rents in 2010 for all West German counties and the regional coverage of the IVD data.

Wage Data. We additionally use linked employer-employee data from the Institute for Employment Research (LIAB) to study the effect of property taxes on wages at the municipal level. The LIAB data is based on a one percent stratified random sample of all German establishments,

[10] We complement this panel with earlier data from the Statistical Offices of the federal states of Bavaria (since 1970), Lower Saxony (since 1981), and Northrhine-Westfalia (since 1977), three of the largest states in Germany.

covering about 15,000 plants. The firm data are linked to social security records, matching all employees working in these plants. Overall, the LIAB data covers between 1.6 and 2.0 million workers per year, which corresponds to about 6 percent of all German workers (see Alda et al., 2005, Fuest et al., 2018, for more information on the dataset). We link the municipal tax information to the LIAB via the plant location, and calculate the mean wage for all municipality-year cells.[11]

1.4.2 Sample Definition

While our panel data set contains the universe of all West German municipalities, the housing data covers only 436 jurisdictions (five percent of all West German municipalities). Panel A of Figure 1.A.3 in the Appendix shows that the housing data set accounts for roughly forty percent of the West German population. Panel B differentiates the sample by city size. The housing data set includes all large cities above 100,000 inhabitants and a substantial part of the medium-sized cities with a population above 20,000. Appendix Figure 1.A.4 shows the size distribution of municipalities in the baseline sample and in the full sample. Despite the difference in population, both samples are rather comparable when looking at the number and the size of property tax changes, i.e., the source of identifying variation in our empirical analysis, as can be seen in Appendix Figure 1.A.5.

In the empirical analysis below, we use the housing data sample as the baseline sample, also for results on other outcomes, in order to have a consistent sample definition.[12] Yet, if data is available, we additionally present estimates on the full sample of all West German municipalities.

We exclude East German municipalities from our analysis for two reasons. First, and foremost, East German housing markets don't seem ideal testing grounds for our theoretical predictions given the tremendous population loss and the large inflow of public and private capital after reunification. In fact, East German municipalities on average lost more than 15 percent of their population since the fall of the Berlin Wall in 1989. As a consequence, housing markets in many East Germany regions have been subject to substantial excess supply during the past decades. Second, there were substantial mergers of East German municipalities after reunification, which complicates any longitudinal study at the municipal level. About 60 percent of the East German municipalities experienced at least one merger since 1990. Given that our data are based on municipal boundaries as of December 31, 2010, we cannot assign the correct tax rates.[13]

[11] About 10 percent of the wages in the LIAB are right censored at the social security contribution ceiling. We get similar results when using the median municipal wage, which should be hardly affected by wage censoring.

[12] Note that for some outcomes we have to restrict the estimation sample even further to only 89 city counties (*kreisfreie Städte*) if data is solely available at the county level.

[13] A possible solution would be to use a (weighted) average of the municipalities which merged, but this would introduce considerable measurement error and several artificial tax changes. Moreover, using these municipalities

1.5 Empirical Model

We make use of an event study design to investigate the effects of property tax changes. As identified by our theoretical model, we are interested in the effects of property taxes on the following outcome variables: net rents, house prices, number of houses, number of apartments, wages, population, land sales, and land prices. Denoting an outcome in municipality c in year t as $y_{c,t}$, our regression model reads as follows:

$$\ln y_{c,t} = \sum_{k=-4}^{6} \beta_k D_{c,t}^k + \mu_c + \zeta_{c,t} + \varepsilon_{c,t}. \tag{1.17}$$

We regress logged outcomes on a set of event study indicators $D_{c,t}^k$ with the event window running from four years prior to a tax reform ($k = -4$) to six years after the event ($k = 6$).

We estimate two different variants, which differ in the way we define event indicators $D_{c,t}^k$. First, we implement the most intuitive and basic model, where $D_{c,t}^k$ is simply a dummy variable indicating a tax increase k years ago.[14] Second, we follow Simon (2016) and Fuest et al. (2018) in estimating a specification where the event indicator $D_{c,t}^k$ switches on only for large tax increases, i.e., increases being equal to or greater than the median of the tax increase distribution. The reason is that tax reforms might only have real effects if tax changes are sizable, for example, due to adjustment costs.

Our baseline specification of the event study includes four leads and six lags, $k = -4, \dots, 6$, which enables us to investigate the dynamics of the relation between property taxes and outcomes on the housing, land, and labor market, where (quantity) responses might take some time (England, 2016).[15] In both models, end points $D_{c,t}^{-4}$ and $D_{c,t}^{6}$ of the event study are adjusted to account for the fact that our panel is unbalanced in event time due to staggered tax reforms across municipalities (McCrary, 2007).[16] This adjustment makes the set of $4 + 6 + 1$ event indicators $D_{c,t}^k$ perfectly collinear and we thus normalize coefficients to the pre-reform year by omitting the respective event indicator $D_{c,t}^{-1}$ from the regression, i.e., $\beta_{-1} = 0$.

in our analysis would require the strong assumption that the decision to merge is unrelated to the fiscal and economic situation in a municipality.

[14] Almost 97 percent of the property tax reforms between 1992–2017 are tax increases. Estimates are not sensitive to whether we keep municipalities with tax decreases in the control group or exclude them from the sample.

[15] Clearly, the choice of the event window is determined by data availability over time. The chosen baseline is a compromise between the length of the event window and statistical power. We experimented with other event window definitions, finding very similar results.

[16] Hence, the coefficient β_{-4} captures the effect of all tax changes occurring four or more years before the current reform. Likewise, the coefficient β_6 measures the effect of all tax changes that happened six or more years after a reform. Since endpoints are estimated on unbalanced data, we follow standard practice and do not plot them in the event study graphs (Smith et al., 2017, Fuest et al., 2018).

To control for time-invariant factors, we include municipality fixed effects μ_c.[17] The vector $\zeta_{c,t}$ controls for local shocks by including state × year fixed effects and linear county trends. Moreover, we include event study coefficients indicating whether the local business tax, the other tax instrument at the disposal of municipalities, changed.[18] The error term is denoted by $\varepsilon_{c,t}$. We allow for clustering of standard errors at the municipal level to account for correlation in unobservable components over time and between the different building and construction types.[19]

Identification of the β_k coefficients comes from changes within a municipality relative to the pre-reform year and relative to the regional trend. The identifying assumption is that there are no other factors that simultaneously affect property taxes and the outcome variables. Using an event study design allows to directly test for reverse causality problems. In order to obtain causal estimates, we need pre-trends to be flat and insignificantly different from zero.

While municipality fixed effects control for time-invariant confounders, our estimator will be biased if local shocks affect both municipal fiscal policies and housing as well as labor markets. We test for confounding factors in two ways. First, we assess the sensitivity of our estimates with respect to the inclusion of a very rich set of time-varying control variables. In our baseline, we include state × year fixed effects and linear county trends. As a robustness check, we estimate one less demanding specification, dropping the linear county trends, and one more demanding one, including commuting zone × year fixed effects. Estimates prove to be robust (see Appendix 1.C.1). Second, we directly test whether tax reforms are driven by the local business cycle by using municipal unemployment and GDP per capita at the county level as outcomes variables in the event study regression in equation (1.17).

1.6 Reduced-Form Results

In this section, we present reduced-form results for the effects of property tax changes on prices and quantities on the housing, land, and labor market. First, we test the theoretical predictions derived from our spatial equilibrium model in Section 1.2. Second, we investigate heterogeneous effects with respect to city size, population density, the availability of developable land, and the housing quality to check whether the mechanisms implied by our theoretical model are supported by the data.

[17] Our data contains several indices for each municipality differing by construction type and building quality. We include all indices and account for type-quality-specific municipality fixed effects in rent and price regressions.

[18] As it turns out, controlling for changes in the local business tax, does not affect our results on the housing or land market. But intuitively, and in line with the results in Fuest et al. (2018), we find differences in wage responses. Ignoring changes in business taxes would lead to more negative wage effects.

[19] The results are not sensitive to whether we cluster at the municipal level or the level of commuting zones.

1.6.1 Main Results

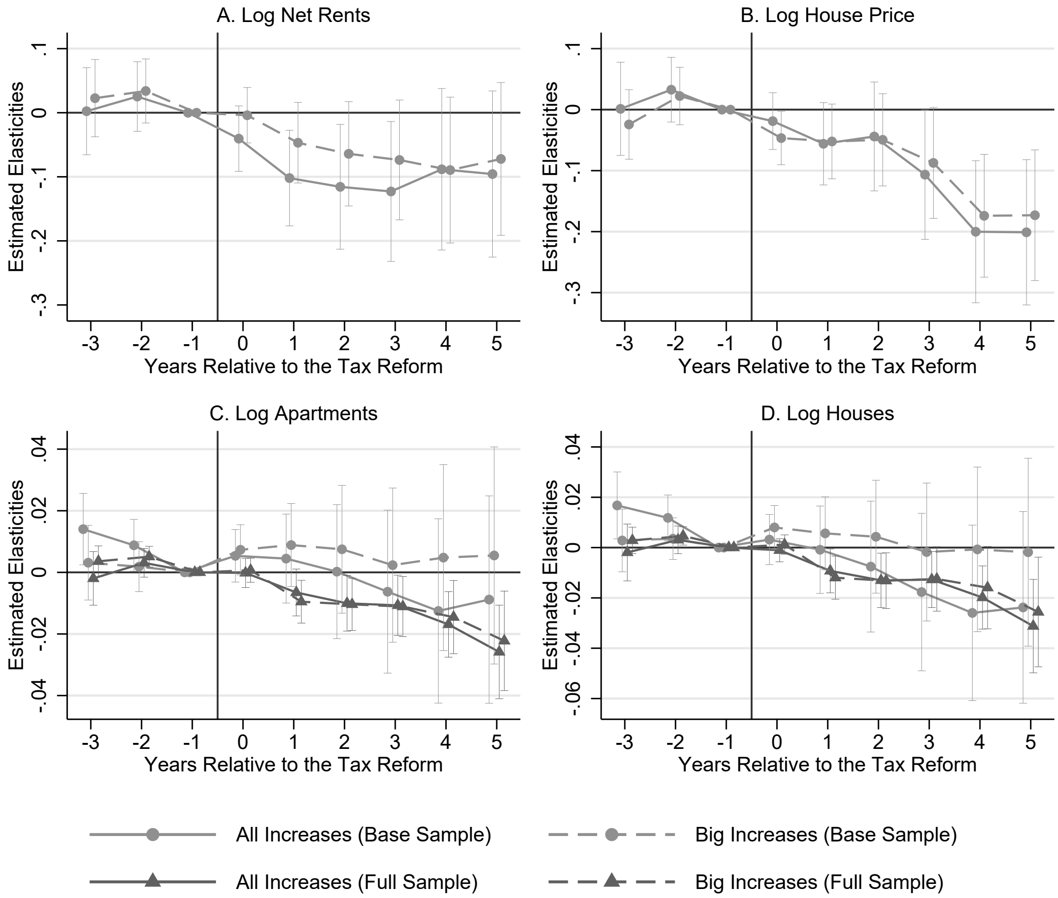
We report event study results for various outcomes on the housing, land, and labor market. For each outcome, we plot two event study specifications: the classic design where event study dummies indicate tax increases, and an alternative specification where event dummies are equal to one for large tax increases only. For both models, estimated treatment effects of (large) tax increases are rescaled such that plotted coefficients can be interpreted as average elasticities. Various sensitivity checks with respect to the length of the event window, and the way we account for local shocks are presented and discussed in Appendix 1.C.1.

Housing Market. We start by analyzing housing market effects, Figure 1.3 summarizes the results. Panel A reports the event study results using log quoted net rents as an outcome. While pre-reform trends are flat as required by our identifying assumption, we find that net rents for new contracts are about 0.1 percent lower in the three years following a tax increase of one percent. This short to medium-run effect is statistically different from zero at the 90 percent level. However, after three years, the negative effect on rents starts to revert slowly towards zero. This implies that in the medium run, more and more of the incidence is borne by the new renter. A likely explanation of this adjustment path lies in the supply of rental dwellings, which is inelastic in the short run but becomes more elastic over time.

Next, we look at the effects of property tax increases on net house prices, which are plotted in Panel B of Figure 1.3. We detect a gradual decrease in house prices after a property tax increase. The implied elasticity five years after the tax increase is about -0.2. Hence, and in line with the literature, we find clear evidence of property tax capitalization into house values. While buyers have to bear the full tax burden in the short run, they are able to shift part of the future tax liability onto the seller of the house, which is reflected in lower transaction prices.

Panels C and D show the quantity effects of rising property taxes on the housing market. While we find a gradually negative effect on the stock of apartments and residential buildings in a municipality, magnitudes are small and results are not statistically significant at conventional levels. The lower magnitudes of the quantity responses are intuitive as the housing stock cannot adjust as quickly as prices. This is also predicted by the theoretical model, where the tax elasticity of floor space equals the tax elasticity of rents multiplied by the effective housing supply elasticity. Because of this sluggish construction response, it is difficult to identify quantity effects that are statistically significantly different from zero. A potential way to overcome this problem is to increase the number of observations. As mentioned in Section 1.4.2, our baseline sample is restricted to municipalities for which we have housing price data. Yet for some variables, we have data from the universe of West German municipalities. While effects might be different across samples due to heterogeneous treatment effects, we can increase statistical

Figure 1.3: The Effects of Property Taxes on the Housing Market



Notes: This figure shows the effects of property taxes on the housing market using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases), or as an increase that is greater than or equal to the median of the tax change distribution (big increases). The base sample includes all municipalities from our housing data set (see Section 1.4.2 for details), the full sample includes all municipalities for which we have data on the respective outcome. Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

power by testing quantity responses in the full sample. For this reason, we plot two additional sets of event studies estimates for the full municipality sample in Panels C and D. We find that the effect on the number of residential houses is very similar across samples, but highly significant in the full sample with an elasticity of around -0.03 six years after a tax increase (Panel D). Moreover, we find an effect of similar magnitude on the number of apartments in the full sample, while effects in the base sample, which covers larger, more urban areas, is closer to zero (Panel C). These findings are in line with evidence from Lutz (2015) who reports capitalization effects in more urban areas, while property taxes reduce residential construction in rural communities. The results again highlight the importance of treatment effect dynamics as the quantity effects take time to evolve, which is again in line with the idea that housing supply is rather inelastic in the short run.

Land Market. Next we turn to the land market. Figure 1.4 summarizes the results. Panel A shows that the land price strongly decreases in response to a tax increase and only gradually starts to recover five years after the tax reform. This is in line with the theoretical prediction from our spatial equilibrium model. Notably, the magnitude of the effect is larger compared to the effects on rents, which suggests that the land supply is relatively inelastic in German municipalities (cf. Section 1.2.6). As above, effects on the quantity of residential land use are not conclusive within the house price sample and we cannot reject the null hypotheses of no effect. However, we find negative and statistically significant effects on the land area used for residential housing in the full sample of municipalities: five years after a tax change, the land use is reduced by 0.07 percent for a one percent tax increase (see Panel B of Figure 1.4).

Figure 1.4: The Effects of Property Taxes on the Land Market

Notes: This figure shows the effects of property taxes on the land market using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases), or as an increase that is greater than or equal to the median of the tax change distribution (big increases). The base sample includes all municipalities from our housing data set (see Section 1.4.2 for details), the full sample includes all municipalities for which we have data on the respective outcome. Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

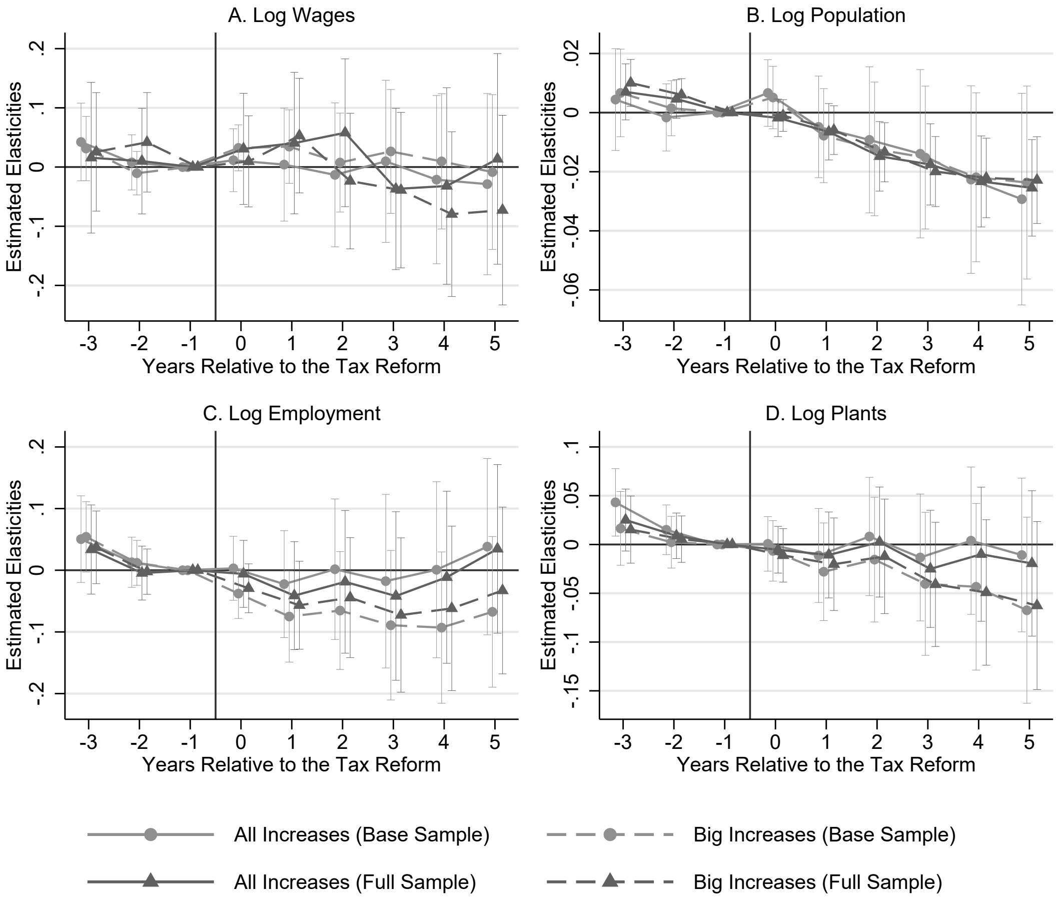
Labor Market. Last, we turn to local labor market outcomes. Panel A of Figure 1.5 shows that wages are largely unaffected by tax increases. In light of our theoretical predictions derived from the spatial equilibrium model, this implies that the rent elasticities of commercial and residential floor space are roughly of similar size. As for the housing stock, we see a gradual decline of around 0.02-0.03 percent in municipal population in response to a one percent tax increase (see Panel B). Effects are again similar in magnitude for the baseline and the full sample, but we again lack statistical power in the house price sample. Panels C and D show a insignificant but visible decline of roughly equal size in local employment and the number of plants in a municipality when we focus on large tax increases, which is in line with the mechanisms predicted by the theoretical model.

Sensitivity Checks. We conduct a wide range of robustness checks. First, we can directly test the identifying assumption that tax reforms are not driven by local business cycles by putting log GDP and log unemployment on the left-hand side of our estimation equation (1.17). As shown in Appendix Figure 1.C.1, pre-trends are very flat for GDP, while we detect a small pre-trend for local unemployment. In terms of post-treatment effects, the figure suggests no effect on local GDP due to the property tax increase. Local unemployment is increasing after tax reforms, in particular in the baseline house price sample. While most effects are not statistically significant, the results suggest that an increase in local property taxes tends to hurt the overall local economy.

Moreover, and in a similar vain, we test the sensitivity of our estimates with respect to confounders. In our baseline specification we include state times year fixed effects and linear county trends. This rich set of non-parametric controls is likely to account for various potentially confounding shocks at the local level. As a sensitivity check, we estimate one more parsimonious specification including only state times year fixed effects but excluding county trends, and one richer specification, where we control for commuting-zone times year fixed effects. There are 204 commuting zones which delineate local labor markets based on commuting flows. Including commuting zone times year fixed effects thus absorbs any common shocks within labor market regions.[20] Figures 1.C.2–1.C.4 in Appendix 1.C.1 show the results. If anything, pre-trends become flatter in the more involved baseline specification. In terms of post-treatment effect, the general pattern is confirmed for most outcomes.

[20] The rich specification is only meaningful for outcomes measures at the municipal level. We are forced to focus on a subset of 88 municipalities (city counties) for the few outcomes that we observe at the county level only, i.e., land prices, employment, number of plants. There are 88 municipalities and 73 commuting zones in this sample, which makes commuting zone × year fixed effects highly collinear with the identifying variation, which is at the municipality-year level.

Figure 1.5: The Effects of Property Taxes on the Labor Market

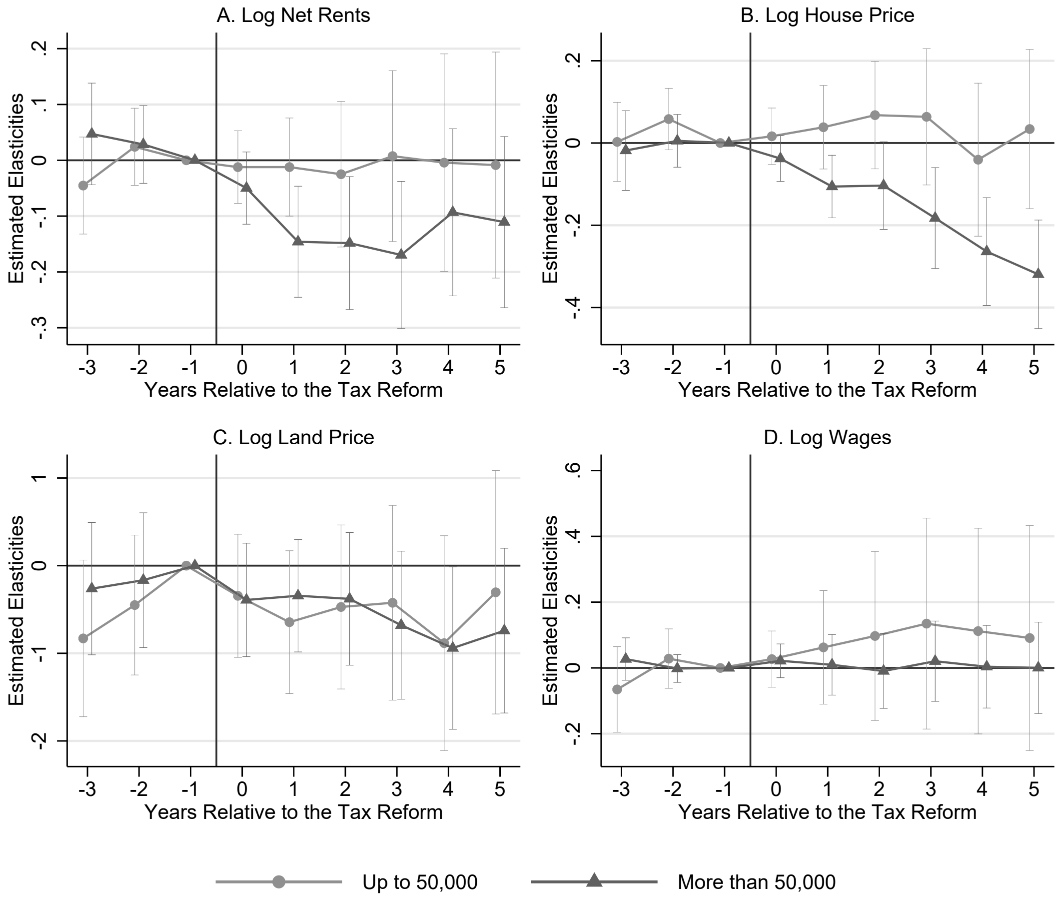


Notes: This figure shows the effects of property taxes on the labor market using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases), or as an increase that is greater than or equal to the median of the tax change distribution (big increases). The base sample includes all municipalities from our housing data set (see Section 1.4.2 for details), the full sample includes all municipalities for which we have data on the respective outcome. Panels C and D are based on a subsample of large municipalities ("urban counties") as data on employment and plants are only observed at the county level. Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

1.6.2 Heterogeneous Effects

We can further test the underlying mechanisms of the theoretical model by estimating heterogeneous effects for certain municipality types. For instance, the theoretical model assumes that the negative effect of property taxes on rents is higher, the less elastic the supply of rental housing. While the adjustment pattern shown in Panel A of Figure 1.3 has provided a first

Figure 1.6: The Effects of Property Taxes on Prices by City Size



Notes: This figure shows the effects of property taxes on prices by city size using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

indication for this mechanism with net rents decreasing in the short run, but reverting to pre-reform levels in the longer run, we can exploit cross-sectional differences in the housing supply elasticity by differentiating between municipalities with below and above 50,000 inhabitants. With strong urbanization trends and growing cities, housing and land supply in cities is rather inelastic. Hence, we would expect to see stronger price reactions in larger municipalities. This is confirmed by Figure 1.6, which shows that rents, house prices and land prices decrease more strongly in larger cities. Conversely, Figure 1.C.10 in Appendix 1.C.2 demonstrates stronger quantity responses in smaller, more rural municipalities.[21] In light of the theoretical predictions, these results suggest that housing and land supply are indeed more elastic in rural areas than in cities.

A different way to investigate heterogeneous effects is to distinguish municipalities by the share of undevelopable land (Hilber and Vermeulen, 2016). Following the rationale of Saiz (2010), municipalities with a high share of undevelopable land are assumed to have less elastic land and housing supply and should see stronger price effects. Figure 1.C.13 in the Appendix confirms this prediction – at least for house and land prices. While the municipalities with a more elastic housing supply hardly face any price changes, they experience stronger quantity responses instead (see Appendix Figure 1.C.14).

Last, we check for differential responses by construction quality, Figure 1.7 shows the results. We find that net rents in apartments of the lowest quality revert back more quickly to the pre-reform level than net rents of higher quality housing (see Panel A). This implies that the pass-through of the tax burden on renters in lower quality housing is higher and faster *ceteris paribus*. Looking at sales prices of houses, Panel B suggests that the tax shifting from buyers to sellers is slightly lower for high quality housing, yet estimates are not statistically different from each other.

1.7 Combining Theory and Empirics

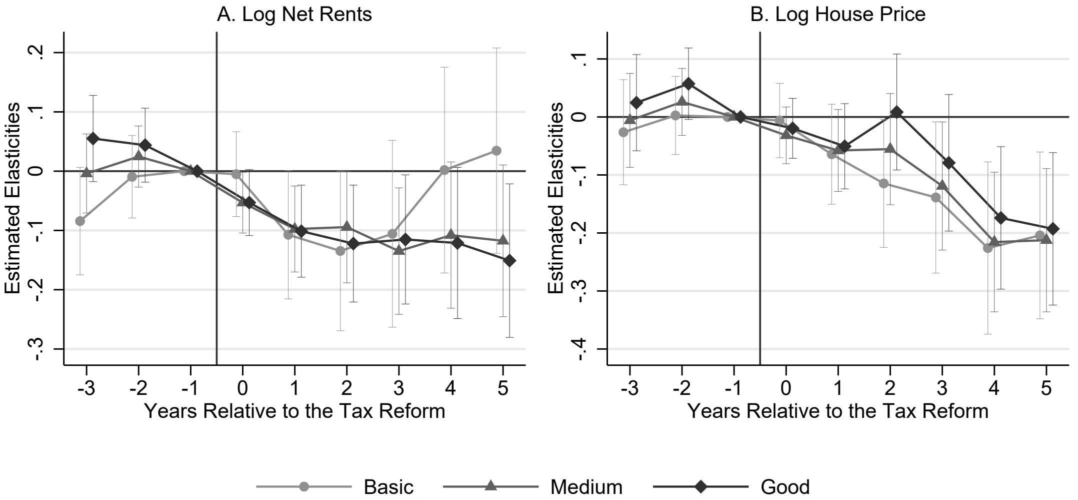
In the following section, we combine the reduced-form estimates presented in Section 1.6.1 with the theoretical model set up in Section 1.2. First, we calculate marginal welfare effects of property tax increases using the reduced form results and the welfare formulas from Section 1.2.7. Second, we assess the role of endogenous amenities in shaping the welfare effects of the property tax by running counterfactual simulations.

1.7.1 Welfare Effects of Property Tax Increases

Ignoring public goods, Proposition 2 shows that the marginal welfare effects of property tax increases are governed by four key elasticities and three additional model parameters, which we calibrate using estimates from the literature and external data sources. In Table 1.1, we summarize the key medium-run elasticities after five years for both the house price sample and

[21] We find similar results when interacting tax hikes with population density indicators (see Figures 1.C.11 and 1.C.12 in Appendix 1.C.2). The heterogeneous responses by city size or density already became apparent when looking at the main results in Section 1.6.1, where quantity responses become stronger and more significant when switching from the mostly urban baseline to the full sample, which also includes smaller and more rural municipalities.

Figure 1.7: The Effects of Property Taxes on Prices by House Quality



Notes: This figure shows the effects of property taxes on prices by house quality using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

the full sample.[22] All elasticities are negative and have the predicted sign. As already discussed above, price effects are larger in magnitude than quantity effects.

Calibrating the housing share in consumption α and the labor share in production β, we can derive the welfare losses of property tax increases for workers, firm owners, and land owners, calculated as respective shares of the marginal welfare effects over the total welfare loss. Figure 1.8 visualizes the results for different assumptions on the calibrated parameters α and β (see Table 1.C.2 in Appendix 1.C.3 for the corresponding numbers). For our preferred baseline values of the housing and labor share, $\alpha = 0.3$ and $\beta = 0.55$, we find that workers bear 22 percent of the welfare loss of property tax increases, firm owners bear 35 percent, and land owners 43 percent.

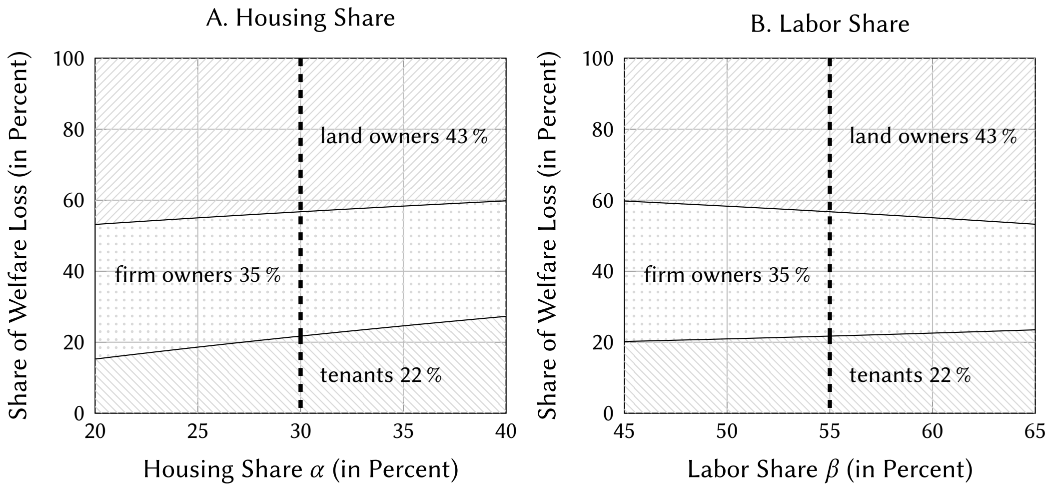
[22] The presented elasticities are the coefficients of the conventional event study design using all tax increases after five years. As noted above, we scale the coefficients such that they represent elasticities, i.e., they measure the effect of the outcome in percent in response to a one percent increase in the local property tax rate. For the simulation, we are interested in the medium to long-run effect of property taxes. The traditional difference-in-difference estimate would provide us with average treatment effect relative to the pre-treatment period. As shown above, most of the effects materialize gradually rather than sharply after treatment, which would mean that the DiD estimate is lower than the medium-term effect. We nevertheless present DiD estimates in Appendix Table 1.C.1 and confirm this pattern empirically. The event study results provides unbiased estimates for medium-run effects in case of flat pre-trends, which is the case for our outcomes.

Table 1.1: Reduced Form Elasticities

	Base Sample			Full Sample		
Outcome	Elasticity	S.E.	Obs.	Elasticity	S.E.	Obs.
Log Net Rent	-0.096	(0.079)	37,672	-0.096	(0.079)	37,672
Log House Price	-0.201	$(0.072)^{***}$	33,767	-0.201	$(0.072)^{***}$	33,767
Log Apartments	-0.009	(0.020)	2,780	-0.026	$(0.009)^{***}$	93,212
Log Houses	-0.024	(0.023)	2,780	-0.031	$(0.011)^{***}$	93,212
Log Land Price	-0.999	(0.687)	1,205	-0.512	(0.524)	6,987
Log Land	-0.004	(0.076)	1,712	-0.072	$(0.030)^{**}$	50,843
Log Wage	-0.029	(0.092)	973	0.014	(0.108)	8,067
Log Population	-0.029	(0.022)	5,050	-0.025	$(0.010)^{**}$	189,535

Notes: This table summarizes the reduced-form results for the key medium-run elasticities of our model for both the house price sample and the full sample. For detailed information on all variables, see Appendix Table 1.A.1.

Figure 1.8: Welfare Effects Without Endogenous Local Public Goods

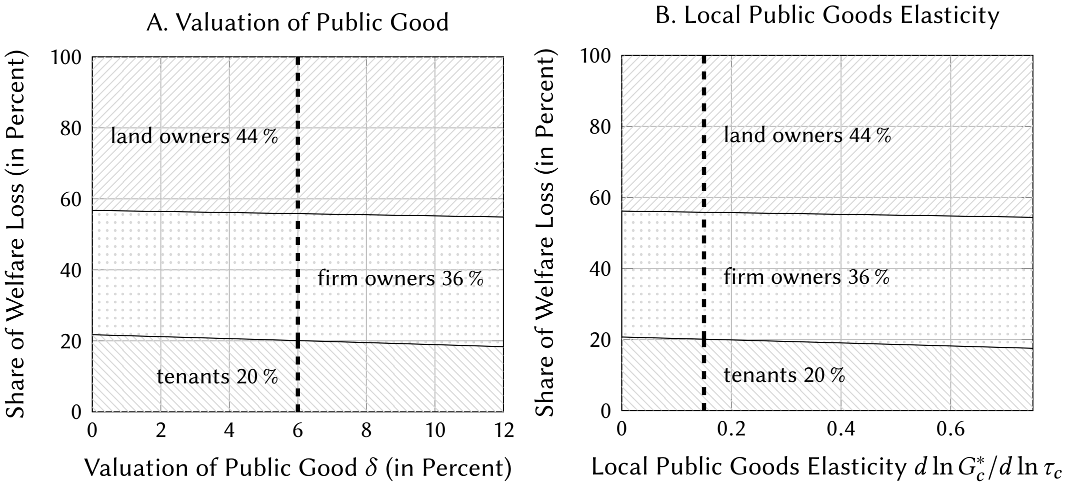


Notes: This figure presents welfare effects arising in a model where property tax revenues are not spent on local amenities. The marginal welfare effects are shown for different values of the housing share in consumption (α) in Panel A (assuming parameter $\beta = 0.55$), and the labor share in the production of the tradable good (β) in Panel B (assuming $\alpha = 0.3$). Marginal welfare effects are based on Proposition 2 in Section 1.2.6 and the following three reduced-form elasticities: $d \ln r_c^{H*}/d \ln \tau_c = -0.096$, $d \ln l_c^*/d \ln \tau_c = -0.512$, and $d \ln w_c^*/d \ln \tau_c = 0.014$ (see full sample results in Table 1.1). The dashed vertical lines mark the baseline specification with parameters calibrated at $\alpha = 0.3$ and $\beta = 0.55$.

1.7.2 The Relevance of Local Public Goods

Next, we assess the empirical relevance of endogenous local public goods. Public goods affect marginal welfare effects in two ways: (i) through the valuation of public relative to private goods, and (ii) via the transmission of property tax increases into additional public good spending. While we do not know the average individual valuation of the public goods, δ, we can calibrate the parameter to the average share of local public expenditures relative to local GDP (Fajgelbaum et al., 2016, Michaillat and Saez, 2017), which yields an average of $\delta = 0.06$. The importance of local public goods for marginal welfare effects also depends on how responsive public good spending is to changes in property taxation. As explained in Section 1.2.6, the direct, negative effects of tax increases on equilibrium prices are alleviated if the tax revenue is spent on local public goods valued by the population (see Proposition 1).

Figure 1.9: Welfare Effects With Endogenous Local Public Goods



Notes: This figure presents welfare effects arising in a model where property tax revenues are partly spent on local amenities. The marginal welfare effects are shown for different assumptions on the valuation of the local public good (δ) in Panel A (assuming parameters $d \ln G_c^* / d \ln \tau_c = 0.15$), and the local public goods elasticity with respect to tax changes ($d \ln G_c^* / d \ln \tau_c$) in Panel B (assuming $\delta = 0.06$). Marginal welfare effects are based on Proposition 2 in Section 1.2.7 and the following three reduced-form elasticities: $d \ln r_c^{H*} / d \ln \tau_c = -0.096$, $d \ln l_c^* / d \ln \tau_c = -0.512$, $d \ln w_c^* / d \ln \tau_c = 0.014$, and parameters $\alpha = 0.3$ and $\beta = 0.55$. The dashed vertical lines mark the baseline specification with parameters calibrated at $\delta = 0.06$ and $d \ln G_c^* / d \ln \tau_c = 0.15$.

Using the reduced form price elasticities and the calibrated parameters α, β, and δ, we can simulate marginal welfare effects over a range of potential elasticities $\frac{d \ln G_c^*}{d \ln \tau_c}$, i.e., different assumptions on the transmission of property tax increase in government spending. Figure 1.9 shows the welfare losses for workers, firm owners, and land owners. Panel A shows the welfare effects for different assumptions on the valuation of public goods δ. Panel B illustrates the welfare impact of property tax increases for different values of the public goods elasticity. Accounting for endogenous local public goods changes the results from Section 1.7.1 only marginally. Even large public good elasticities with respect to tax increases hardly lower the welfare loss for tenants within a reasonable range of assumptions for the valuation of public goods. When considering that taxes are used to finance local public goods, tenants bear around 20 percent of the welfare loss of higher taxes, firm owners roughly 36, and land owners 44 percent.

1.8 Conclusion

We propose a new theoretical angle to study the incidence and welfare implications of property taxation by introducing property taxes into a local labor market in the spirit of Moretti (2011). Our spatial equilibrium models encompasses elements of both the capital and the benefit view of property taxation. It also nests simple partial-equilibrium analyses of the housing market. We show that rising local property taxes should lead to a decrease in housing and land prices, a decrease in the housing stock and the use of developed land, and a decrease in municipal population levels. Based on the price effects of property tax increases, the model also allows us to quantify marginal welfare effects.

In the second part of the paper, we empirically test the theoretical predictions of the model using rich administrative panel data and the institutional setting of German municipalities. An event study research design enables us to study treatment effect dynamics and to test for reverse causality in a straightforward manner. We confirm our theoretical predictions. In particular, we show that house prices, the housing and apartment stock as well as population levels decrease if property taxes increase in a municipality. We find no evidence of compensating wage increases following a tax hike.

The results for the net housing prices (rents and house price) inform about the incidence of the property tax. While we detect a short-run decrease of net rents following a tax increase, net rents start to revert back to the pre-reform level three years after the tax reform. This implies that in the medium-run the economic incidence of the property tax is largely on renters. House prices show a negative response, which is still visible after five years. Hence, house buyers are able to shift part of their future property tax burden on the sellers (which are either previous owners or construction companies for newly built houses). Using the reduced-form results we also quantify the welfare implications of property tax increases for the different agents in our model. Our simulations show that workers and tenants bear roughly 20 percent of the welfare loss, firm owners 36 percent, and land owners the largest share with 44 percent of the welfare loss. Importantly, the simulated welfare effects change only little when accounting for the fact that property taxes may be used to finance public goods and local public good provision may thus be endogenous.

Appendix 1.A Data Appendix

Table 1.A.1: Definition of Variables and Data Sources

Variable	Years	Source
Property Tax Rates	1970–2017	The local property tax rate is calculated as the product of local property tax multipliers and an average federal tax rate of 0.32 percent. Data on property tax multipliers is provided by the Federal Statistical Office and the Statistical Offices of the federal states. Data for the period 1998–2017 are published by the Statistical Offices (*Hebesätze der Realsteuern*), data prior to 1998 were provided by the Statistical Offices of the federal states and are partly available online.
Net Rents	1972–2013	Data on quoted net rents are provided as indices for different market segments in the *IVD-Wohn-Preisspiegel* by the German real estate association IVD (*Immobilienverband Deutschland*). We validate these indices against county-level data from *empirica Preisdatenbank* for 2004–2013, *bulwiengesa AG (RIWIS)* for 1990–2013, and the German micro census provided by the RDC of the Federal Statistical Office and Statistical Offices of the Länder (1998, 2002, 2004, 2010, 2014).
House Prices	1972–2013	Data on quoted house prices are provided by the German real estate association IVD in the *IVD-Wohn-Preisspiegel* in addition to data on net rents.
Housing Stock	2001–2015	Data on the number of apartments (*Wohnungen in Wohn- und Nichtwohngebäuden*) and the number of residential buildings (*Wohngebäude*) in a municipality are provided by the Federal Statistical Office and the Statistical Offices of the federal states. Data for the years 2001–2007 stem from the database *Statistik lokal*, data for later years are published online in the database *Regionalstatistik*.
Wages	1999–2008	Data on wages as of June 30 are available in the linked employer-employee data of the Institute for Employment Research (LIAB, see Alda et al., 2005). The data set is based on a 1 % stratified random sample of all German establishments and contains information on all employees working in these plants. Sampling is based on the location of the establishment not the residence of the worker. Social security data covers more than 80 percent of the work force in Germany but excludes civil servants and self-employed individuals.
	1996–2009 (counties)	Data on yearly wages in German counties are provided by the Working Group Regional Accounts (*Volkswirtschaftliche Gesamtrechnung der Länder, Revision 2005*). We restrict the sample to city counties (*kreisfreie Städte*) and discard rural counties that contain more than one municipality to avoid measurement error.

continued

Table 1.A.1 continued

Variable	Years	Source
Population	1970–2017	Data on municipal population are provided by the Federal Statistical Office and the Statistical Offices of the federal states combined with local property tax multipliers.
Employment	1996–2009 (counties)	Data on the average number of employed individuals in a county come from the official employment statistics of the Federal Statistical Office and the Statistical Offices of the federal states. We restrict the sample to city counties (*kreisfreie Städte*) and discard rural counties that contain more than one municipality to avoid measurement error.
Plants	1999–2012 (counties)	Yearly data on the number of establishments in a county come from the Federal Employment Agency (*Bundesagentur für Arbeit*). We restrict the sample to city counties (*kreisfreie Städte*) and discard rural counties that contain more than one municipality to avoid measurement error.
Land Prices	1995–2013 (counties)	Data on land transaction prices are provided online in the database *Regionalstatistik* for all German counties by the Federal Statistical Office and the Statistical Offices of the federal states. We restrict the sample to city counties (*kreisfreie Städte*) and discard rural counties that contain more than one municipality to avoid measurement error.
Land Use	2008–2015	Data on land use in German municipalities is provided online in the database *Regionalstatistik* by the Federal Statistical Office and the Statistical Offices of the federal states. We use the area covered by buildings and the surrounding land assigned to residential or commercial use (*Gebäude- und Freifläche*, measured in hectare) as our indicator of land used for construction.
Spending/Revenues	1998–2008	Data on municipal revenues and municipal spending are provided by the Federal Statistical Office and the Statistical Offices of the federal states. Data for the years 2001–2008 stem from the database *Statistik lokal*, data prior to 2001 come from the Statistical Offices of the federal states.
Local GDP	1992–2009 (counties)	Data on the gross domestic product in a German county is provided by the Working Group Regional Accounts (*Volkswirtschaftliche Gesamtrechnung der Länder, Revision 2005*). We restrict the sample to city counties (*kreisfreie Städte*) and discard rural counties that contain more than one municipality to avoid measurement error.
Unemployment	1998–2013	The number of unemployed individuals in a municipality is published by the Federal Employment Agency (*Bundesagentur für Arbeit, Arbeitslose nach Gemeinden*).

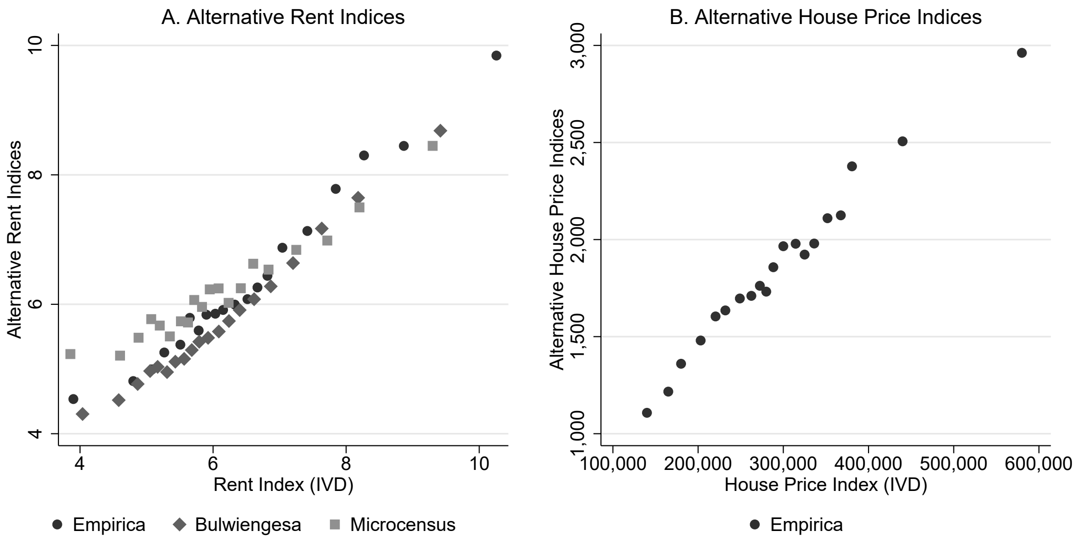
Notes: This table summarizes the definition of the used variables and provides details on the data sources.

Table 1.A.2: Descriptive Statistics

	Mean	SD	P25	P50	P75	Min	Max	N
Panel A – Price Data Sample								
Local Property Tax Rate	1.20	0.22	1.04	1.18	1.33	0.26	2.49	5,593
Number of Tax Changes 1992–2017	5.88	2.49	4.00	6.00	8.00	0.00	15.00	5,593
Change in Property Tax Rate	0.02	0.06	0.00	0.00	0.00	-0.51	0.94	5,531
Log Population	11.06	1.17	10.34	10.98	11.74	6.25	14.41	5,593
Log Houses	9.34	0.93	8.85	9.36	9.85	6.54	12.40	2,840
Log Apartments	10.18	1.12	9.47	10.12	10.87	7.20	13.72	2,840
Log Wage	3.11	0.08	3.05	3.10	3.14	2.93	3.48	840
Log Land Price	5.11	0.70	4.65	5.15	5.60	1.53	7.29	1,296
Log Land Sales	3.62	1.19	2.89	3.74	4.43	0.00	6.93	1,391
Log District Plants	8.36	0.80	7.75	8.26	8.84	6.85	10.85	1,054
Log Employment	9.90	1.28	9.16	9.88	10.71	5.25	13.66	3,969
Log Net Rent	1.73	0.31	1.53	1.74	1.94	0.57	3.77	41,779
Log House Price	12.31	0.50	11.99	12.29	12.61	4.59	14.36	37,072
Log Flat Price	7.30	0.45	7.02	7.35	7.60	4.61	9.39	27,858
Panel B – Full Sample								
Local Property Tax Rate	0.98	0.19	0.86	0.96	1.08	0.00	3.06	277,182
Number of Tax Changes 1992–2017	5.03	2.55	3.00	5.00	6.00	0.00	19.00	277,182
Change in Property Tax Rate	0.01	0.05	0.00	0.00	0.00	-3.06	2.04	268,769
Log Population	7.64	1.40	6.75	7.60	8.60	1.10	14.40	277,182
Log Houses	6.32	1.32	5.45	6.31	7.27	0.00	11.37	90,383
Log Apartments	6.70	1.43	5.72	6.66	7.73	0.00	12.64	90,383

Notes: This table presents descriptive statistics on the variables used. For detailed information on all variables, see Appendix Table 1.A.1.

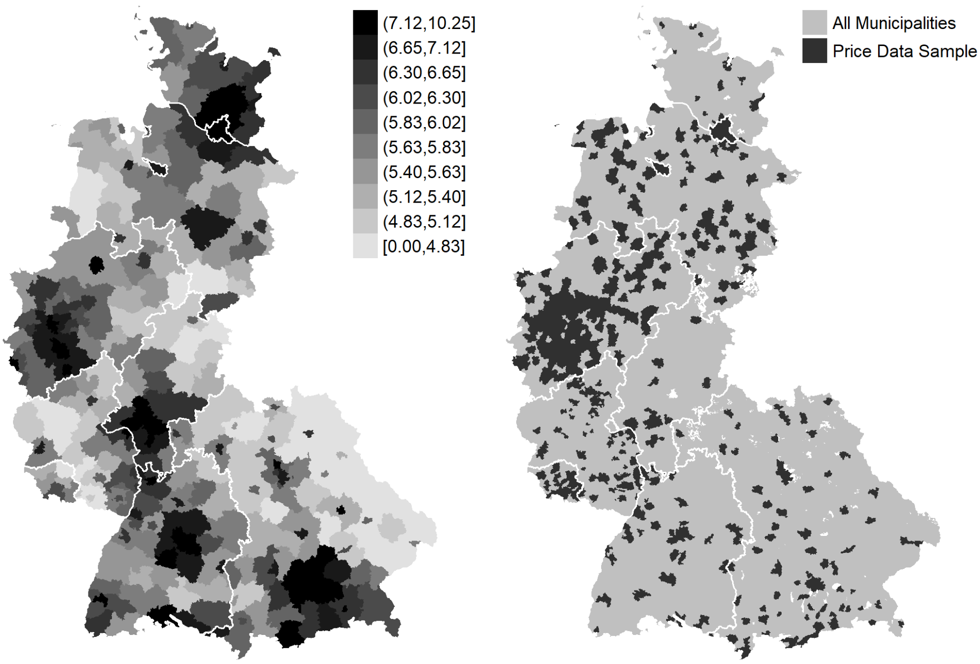
Figure 1.A.1: Alternative Rent and House Price Measures



Notes: Panel A shows binned scatter plots for alternative rent price indices relative to our baseline IVD data. Panel B plots binned scatter plots for an alternative house price index, empirica prices are measured per square meter, IVD prices refer to total prices. Rents and prices measured in EUR. *Sources:* IVD-Wohn-Preisspiegel 1970–2013; empirica Preisdatenbank; bulwiengesa AG, RIWIS; RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 1998–2010, own calculations.

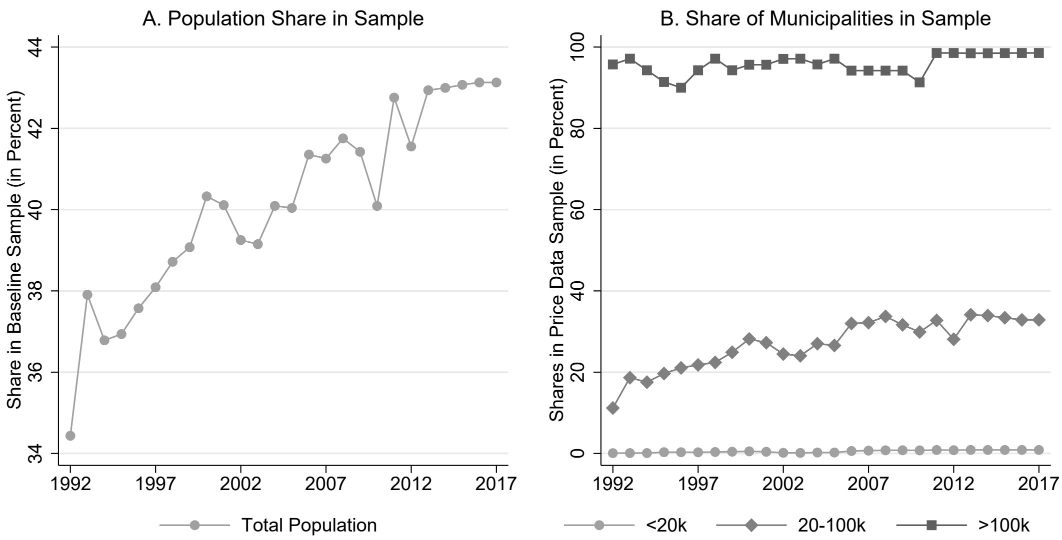
Figure 1.A.2: Rents and Housing Price Data in West Germany

(a) Average County-Level Rents in 2010 (in EUR) (b) Regional Spread of Price Data Sample



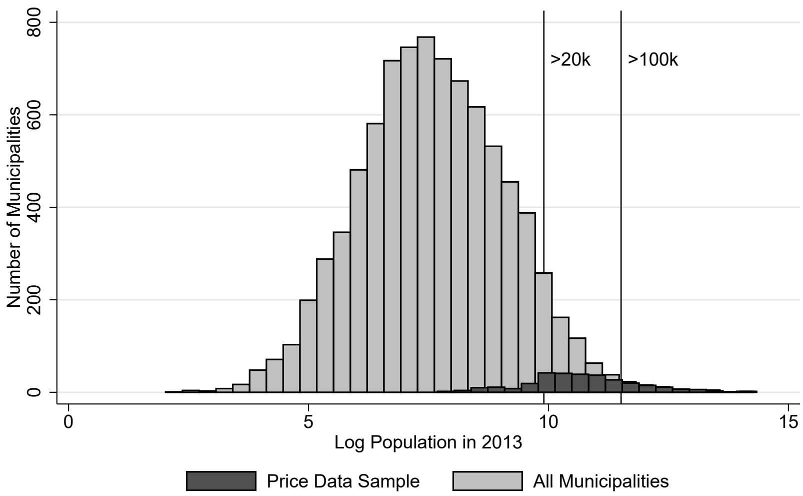
Notes: The left panel of this figure shows average residential rents in 2010 at the county level (measured in 2010 EUR). The right panel shows the geographical distribution of the base sample for which we have house price data. Jurisdictional boundaries are as of December 31, 2010. White lines indicate federal state borders. *Sources:* RDC of the Federal Statistical Office and Statistical Offices of the Länder, Microcensus, 2010, own calculations; IVD-Wohn-Preisspiegel 1970–2013. *Maps:* © GeoBasis-DE / BKG 2015.

Figure 1.A.3: Share of Population and Municipalities in Price Data Sample (in Percent)



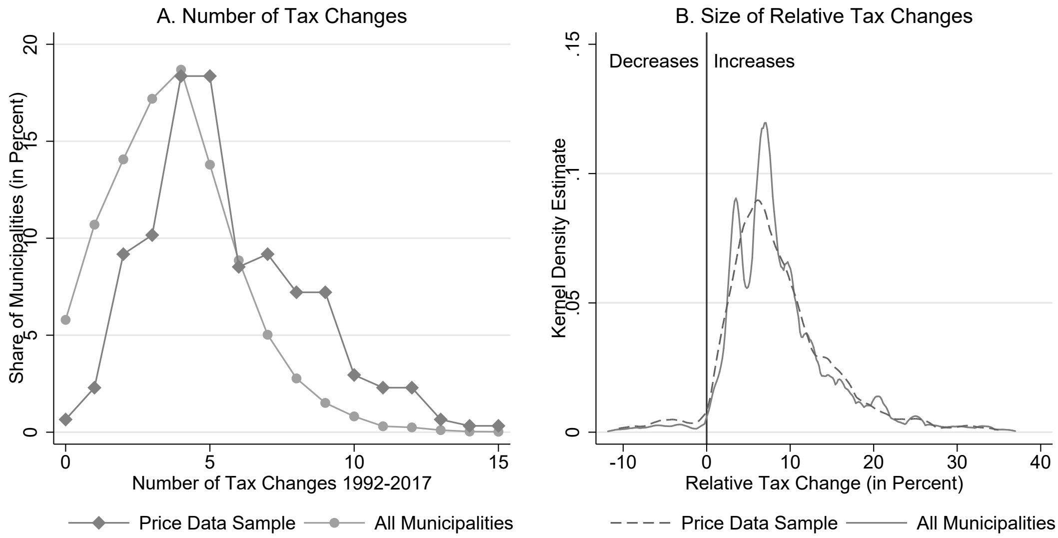
Notes: Panel A plots the share of the West German population that is included in our price data IVD estimation sample over time. Panel B shows the share of West German municipalities that is included in the price data estimation sample over time and by municipality size.

Figure 1.A.4: Number of Municipalities by Size



Notes: This figure plots the size distribution of all municipalities in West Germany in light grey and the size distribution of municipalities in our price data estimation sample in dark grey. Size is measured as log population in 2013. Vertical lines indicate the population thresholds of 20,000 and 100,000 inhabitants. For details on all variables see Appendix Table 1.A.1.

Figure 1.A.5: Number and Size of Tax Changes



Notes: Panel A plots the distribution of the number of tax changes in the period 1992–2017 for all West German municipalities and for the municipalities in our price data sample. Panel B shows the kernel density estimate of the size distribution of relative tax changes for both groups of municipalities (excluding zeros and truncated at the bottom and top one percent). For details on all variables see Appendix Table 1.A.1.

Appendix 1.B Theoretical Model

In this appendix, we provide an extended and more detailed description of our spatial equilibrium model outlined in Section 1.2. Most importantly, it includes all derivations and intermediate steps needed to solve the model and analyze the equilibrium properties. The appendix is self-contained and consequently reiterates and replicates parts of Section 1.2.

We introduce local property taxation into a Rosen-Roback type general equilibrium model of local labor markets (Moretti, 2011). There are $N = 1$ workers that locate in one out of C cities. The model consists of four groups of agents: workers, firms, construction companies and land owners. We solve for the spatial equilibrium and use comparative statics to show how changes in the property tax rate affect the equilibrium outcomes, i.e., population size, floor space, land use, rents, wages and land prices.

1.B.1 Workers

We assume that labor is homogeneous and each worker, i, provides inelastically one unit of labor. Each worker earns a wage w_c and pays rent r_c^H for residential floor space.[23] Each municipality c has a specific unproductive consumption amenity A_c that is exogenously given. In addition, there are endogenous local public goods G_c provided by the local government. Workers maximize utility over floor space h_i, a composite good bundle x_i of non-housing goods and locations c. We normalize the aggregate price of the composite good bundle to one. Workers are mobile across municipalities, but mobility is imperfect due to individual location preferences, so that local labor supply is not necessarily infinitely elastic. In addition to the net house price, there is a property tax in each city, denoted by t_c, with the statutory incidence on the user of the housing service.[24] We assume that households have preferences for public goods measured by $\delta \in (0, 1)$.

The household's maximization problem in a given municipality c is given by:

$$\max_{h_i, x_i} U_{ic} = A_c G_c^{\delta} \left(h_i^{\alpha} x_i^{1-\alpha}\right)^{1-\delta} e_{ic} \qquad \text{s.t. } r_c^H(1 + t_c)h_i + px_i = w_c \tag{1.B.1}$$

with the bundle x_i of non-housing goods Z and the normalized aggregate price index p defined as in Melitz (2003):

$$x_i = \left(\int_{z \in Z} x_{iz}^{\rho} \, \mathrm{d}z\right)^{\frac{1}{\rho}} \qquad\qquad p = \left(\int_{z \in Z} p_{iz}^{-\frac{\rho}{1-\rho}} \, \mathrm{d}z\right)^{-\frac{1-\rho}{\rho}} = 1 \tag{1.B.2}$$

[23] We assume that there is only one homogeneous housing good and do not differentiate between owner-occupied and rental housing in our model (Poterba, 1984).

[24] For simplicity, we assume that property is taxed *ad valorem*. Our main theoretical prediction regarding the tax incidence is however unchanged when modeling the property tax as a specific tax instead (see Appendix 1.B.9).

and $h_i, x_{iz}, A_c, G_c, r_c^H, w_c, t_c, p_{iz}, e_{ic} > 0$ and $\alpha, \rho \in (0,1)$. The parameter ρ relates to the elasticity of substitution between any two composite goods, which is given by $\frac{1}{1-\rho}$ (Dixit and Stiglitz, 1977). The Lagrangian reads:

$$\begin{aligned}\max_{h_i, x_i} \mathcal{L} = &\ln A_c + \delta \ln G_c + \alpha(1-\delta)h_i + (1-\alpha)(1-\delta)\ln x_i + \ln e_{ic} \\ &+ \lambda\left(w_c - r_c^H[1+t_c]h_i - x_i\right)\end{aligned} \tag{1.B.3}$$

and first-order conditions of the household problem are given by:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial h_i} &= \frac{\alpha(1-\delta)}{h_i} - \lambda r_c^H(1+t_c) \overset{!}{=} 0 \\ \frac{\partial \mathcal{L}}{\partial x_i} &= \frac{(1-\alpha)(1-\delta)}{x_i} - \lambda \overset{!}{=} 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= w_c - r_c^H(1+t_c)h_i - x_i \overset{!}{=} 0\end{aligned}$$

Now we can solve by substitution. The optimal floor space consumption is then given by:

$$\begin{aligned}\frac{\alpha(1-\delta)}{h_i} &= \lambda r_c^H(1+t_c) \\ &= \frac{(1-\alpha)(1-\delta)}{x_i} r_c^H(1+t_c) \\ h_i &= \frac{\alpha}{1-\alpha}\frac{x_i}{r_c^H(1+t_c)} \\ &= \frac{\alpha}{1-\alpha}\frac{w_c - r_c^H(1+t_c)h_i}{r_c^H(1+t_c)} \\ &= \frac{\alpha}{1-\alpha}\left(\frac{w_c}{r_c^H(1+t_c)} - h_i\right) \\ h_i^* &= \alpha\frac{w_c}{r_c^H(1+t_c)}\end{aligned} \tag{1.B.4}$$

and we can solve for the optimal consumption level of the composite good bundle:

$$\begin{aligned}x_i &= w_c - r_c^H(1+t_c)h_i \\ &= w_c - r_c^H(1+t_c)\alpha\frac{w_c}{r_c^H(1+t_c)} \\ x_i^* &= (1-\alpha)w_c\end{aligned} \tag{1.B.5}$$

where α is the share of the household's budget spent for housing. Household i's demand of good variety z is then given by $x_{iz}^* = (1-\alpha)w_c p_{iz}^{-\frac{1}{1-\rho}}$. Using the optimal consumption quantities, log indirect utility is defined as:

$$
\begin{aligned}
V_{ic}^H &= \ln U(h_i^*, x_i^*, A_c, G_c, e_{ic}) \\
&= \alpha(1-\delta)\ln h_i^* + (1-\alpha)(1-\delta)\ln x_i^* + \ln A_c + \delta \ln G_c + \ln e_{ic} \\
&= \alpha(1-\delta)\ln\left(\alpha \frac{w_c}{r_c^H(1+t_c)}\right) + (1-\alpha)(1-\delta)\ln([1-\alpha]w_c) + \ln A_c + \delta\ln G_c + \ln e_{ic} \\
&= \underbrace{(1-\delta)\left(\alpha\ln\alpha + [1-\alpha]\ln[1-\alpha]\right)}_{=a_0} + \ln A_c + \delta \ln G_c + \ln e_{ic} \\
&\quad + (1-\delta)(\alpha\ln w_c - \alpha\ln r_c^H - \alpha\ln[1+t_c] + (1-\alpha)\ln w_c) \\
V_{ic}^H &= a_0 + \underbrace{(1-\delta)(\ln w_c - \alpha\ln r_c^H - \alpha\ln[1+t_c]) + \ln A_c + \delta\ln G_c}_{=V_c^H} + \ln e_{ic}.
\end{aligned}
$$

We defined a constant term $a_0 = (1-\delta)(\alpha\ln\alpha + [1-\alpha]\ln[1-\alpha])$ that is the same for all workers in the economy to simplify the notation. The individual (indirect) utility is a combination of this constant a_0, a common term V_c^H identical to all workers in the municipality and the idiosyncratic location preferences e_{ic}. As in Kline and Moretti (2014), we assume that the logarithm of e_{ic} is independent and identically extreme value type I distributed with scale parameter $\sigma^H > 0$. The corresponding cumulative distribution function is $F(z) = \exp\left(-\exp[-z/\sigma^H]\right)$. Due to these city preferences, workers are not fully mobile between cities and real wages $\frac{w_c}{r_c^H(1+t_c)}$ do not fully compensate for different amenity levels A_c across municipalities (other than in Brueckner, 1981). The greater σ^H, the stronger workers' preference for given locations and the lower workers' mobility. There is a city-worker match that creates a positive rent for the worker and decreases mobility. A worker i will prefer municipality a over municipality b if and only if:

$$
\begin{aligned}
V_{ia}^H &\geq V_{ib}^H \\
V_a^H + \ln e_{ia} &\geq V_b^H + \ln e_{ib} \\
V_a^H - V_b^H &\geq \ln e_{ib} - \ln e_{ia}.
\end{aligned}
$$

Given the distribution of $\ln e_{ic}$, it follows that the difference in preferences between two municipalities follows a logistic distribution with scale parameter σ^H, i.e., $\ln e_{ib} - \ln e_{ia} \sim logistic(0, \sigma^H)$. The probability that worker i locates in municipality c when choosing between C cities is then:

$$
N_c = \Pr\left(V_{ic}^H \geq V_{ij}^H, \forall j \neq c\right) = \frac{\exp\left(V_c^H/\sigma^H\right)}{\sum_{k=1}^{C}\exp\left(V_k^H/\sigma^H\right)}.
$$

This expression is equivalent to the share of workers locating in municipality c given that we normalize the total number of workers N to one. Note that the term a_0 cancels out as it is constant across municipalities. Taking logs we arrive at the (log) labor supply curve in municipality c:

$$\ln N_c^S = \frac{V_c^H}{\sigma^H} \underbrace{- \ln\left(C\pi^H\right)}_{=a_1}$$

$$\ln N_c^S = \underbrace{\frac{1-\delta}{\sigma^H}}_{=\epsilon^{\mathrm{NS}}} \ln w_c \underbrace{- \frac{\alpha(1-\delta)}{\sigma^H}}_{=1+\epsilon^{\mathrm{HD}}} \ln r_c^H \underbrace{- \frac{\alpha(1-\delta)}{\sigma^H}}_{=1+\epsilon^{\mathrm{HD}}} \ln \tau_c + \underbrace{\frac{1}{\sigma^H}}_{=\epsilon^{\mathrm{A}}} \ln A_c + \frac{\delta}{\sigma^H} \ln G_c + a_1 \qquad (1.B.6)$$

where we define all terms constant across municipalities as $a_1 = -\ln\left(C\pi^H\right)$ with $\pi^H = \frac{1}{C}\sum_{k=1}^{C} \exp\left(V_k^H/\sigma^H\right)$ being the average utility across all municipalities and we rewrite the property tax rate as $\tau_c = 1 + t_c$. Note that C is given and for large C, a change in V_c^H does not affect the average utility π^H. The labor supply elasticity is given by:

$$\frac{\partial \ln N_c^S}{\partial \ln w_c} = \frac{1-\delta}{\sigma^H} = \epsilon^{\mathrm{NS}} > 0. \qquad (1.B.7)$$

Floor Space Demand. Demand for residential housing in city c is determined by the number of workers in city c and their individual housing demand as indicated by equation (1.B.4):

$$H_c = N_c h_i^* = N_c \alpha \frac{w_c}{r_c^H(1+t_c)}$$

$$\ln H_c = \ln N_c + \ln \alpha + \ln w_c - \ln r_c^H - \ln \tau_c. \qquad (1.B.8)$$

It follows that the intensive margin housing demand elasticity conditional on location choice is equal to -1. In addition, there is an extensive margin with people leaving the city in response to higher costs of living. The aggregate residential housing demand elasticity is given by:

$$\frac{\partial \ln H_c}{\partial \ln r_c^H} = \frac{\partial \ln N_c}{\partial \ln r_c^H} - 1 = -\frac{\alpha(1-\delta) + \sigma^H}{\sigma^H} = \epsilon^{\mathrm{HD}} < 0.$$

1.B.2 Firms

Firms $j = 1, \ldots, J$ are monopolistically competitive and produce tradable consumption goods. Each firm produces a different variety Y_{jc} using labor N_{jc} and commercial floor space M_{jc}. Firms have different productivity across places, due to exogenous local production amenities measured by B_c and idiosyncratic productivity shifters ω_{jc}. Firm j's profits in city c are then given by:

$$\begin{aligned} \Pi_{jc}^{F} &= p_{jc}Y_{jc} - w_c N_{jc} - r_c^M(1+t_c)\kappa M_{jc} \\ Y_{jc} &= B_c \omega_{jc} N_{jc}^{\beta} M_{jc}^{1-\beta} \end{aligned} \tag{1.B.9}$$

with $Y_{jc}, N_{jc}, M_{jc}, p_{jc}, w_c, r_c^M > 0$. w_c and r_c^M denote the factor prices of labor and commercial floor space, respectively. The scale parameter $\kappa > 0$ allows property taxes on commercial rents to differ from residential property taxes. Following Melitz (2003), we substitute the final good price p_{jc} by the inverse of product j's aggregate demand function:

$$Y_{jc} = Q\left(\frac{p_{jc}}{p}\right)^{-\frac{1}{1-\rho}}$$

with price index $p = 1$ as normalized above and $Q > 0$ as total product demand in the economy. The parameter ρ relates to the elasticity of substitution between any two varieties. We define the exponent $-\frac{1}{1-\rho}$ as the constant product demand elasticity $\epsilon^{\text{PD}} < -1$. We can rewrite firm j's profits as:

$$\Pi_{jc}^{F} = \underbrace{Q^{1-\rho} Y_{jc}^{-(1-\rho)}}_{=p_{jc}} Y_{jc} - w_c N_{jc} - r_c^M(1+t_c)\kappa M_{jc}.$$

Using the production function for Y_{jc} we can rewrite this expression as:

$$\Pi_{jc}^{F} = Q^{1-\rho}\underbrace{\left(B_c \omega_{jc} N_{jc}^{\beta} M_{jc}^{1-\beta}\right)}_{=Y_{jc}}^{\rho} - w_c N_{jc} - r_c^M(1+t_c)\kappa M_{jc} \tag{1.B.10}$$

with $B_c, \omega_{jc} > 0$ and $\beta \in (0,1)$.

Profit maximizing behavior leads to the following first-order conditions for labor and floor space:

$$\begin{aligned} \frac{\partial \Pi_{jc}^{F}}{\partial N_{jc}} &= \rho\beta Q^{1-\rho} B_c^{\rho} \omega_{jc}^{\rho} N_{jc}^{\rho\beta-1} M_{jc}^{\rho(1-\beta)} - w_c \overset{!}{=} 0 \\ \frac{\partial \Pi_{jc}^{F}}{\partial M_{jc}} &= \rho(1-\beta) Q^{1-\rho} B_c^{\rho} \omega_{jc}^{\rho} N_{jc}^{\rho\beta} M_{jc}^{\rho(1-\beta)-1} - r_c^M(1+t_c)\kappa \overset{!}{=} 0. \end{aligned}$$

Again, we shorten notation by using $\tau_c = (1 + t_c)$. Taking logs of the second condition we can derive the floor space demand of firms conditional on labor input, factor prices and local productivity:

$$
\begin{aligned}
\ln\left(r_c^M \tau_c \kappa\right) &= \ln\rho + \ln(1-\beta) + (1-\rho)\ln Q + \rho\ln B_c + \rho\ln\omega_{jc} + \rho\beta\ln N_{jc} \\
&\quad - (1-\rho[1-\beta])\ln M_{jc} \\
\ln M_{jc} &= \Big(\ln\rho + \ln[1-\beta] + [1-\rho]\ln Q + \rho\ln B_c + \rho\ln\omega_{jc} + \rho\beta\ln N_{jc} \\
&\quad - \ln r_c^M - \ln[\tau_c\kappa]\Big)\Big/\Big(1-\rho[1-\beta]\Big).
\end{aligned} \tag{1.B.11}
$$

We can derive log labor demand from the first first-order condition using the conditional factor demand for commercial floor space from equation (1.B.11):

$$
\begin{aligned}
\ln w_c &= \ln\rho + \ln\beta + (1-\rho)\ln Q + \rho\ln B_c + \rho\ln\omega_{jc} - (1-\rho\beta)\ln N_{jc} + \rho(1-\beta)\ln M_{jc} \\
\ln N_c &= \Big(\ln\rho + \ln\beta + [1-\rho]\ln Q + \rho\ln B_c + \rho\ln\omega_{jc} + \rho(1-\beta)\ln M_{jc} - \ln w_c\Big)\Big/\Big(1-\rho\beta\Big) \\
&= \Big(\ln\rho + \ln\beta + [1-\rho]\ln Q + \rho\ln B_c + \rho\ln\omega_{jc} \\
&\quad + \rho[1-\beta]\Big[\ln\rho + \ln\{1-\beta\} + \{1-\rho\}\ln Q + \rho\ln B_c + \rho\ln\omega_{jc} + \rho\beta\ln N_{jc} \\
&\quad - \ln r_c^M - \ln\{\tau_c\kappa\}\Big]\Big/\Big[1-\rho\{1-\beta\}\Big] - \ln w_c\Big)\Big/\Big(1-\rho\beta\Big) \\
\ln N_{jc}^* &= \Big(\ln\rho + [1-\rho+\rho\beta]\ln\beta + \rho[1-\beta]\ln[1-\beta] + [1-\rho]\ln Q + \rho\ln B_c \\
&\quad - [1-\rho+\rho\beta]\ln w_c - \rho[1-\beta]\ln r_c^M - \rho[1-\beta]\ln[\tau_c\kappa] + \rho\ln\omega_{jc}\Big)\Big/\Big(1-\rho\Big)
\end{aligned} \tag{1.B.12}
$$

Using equation (1.B.11) from above and firm j's labor demand in city c we can also solve for the commercial floor space demand of firm j:

$$
\begin{aligned}
\ln M_{jc}^* &= \Big(\ln\rho + \rho\beta\ln\beta + [1-\rho\beta]\ln[1-\beta] + [1-\rho]\ln Q + \rho\ln B_c - \rho\beta\ln w_c \\
&\quad - [1-\rho\beta]\ln r_c^M - [1-\rho\beta]\ln[\tau_c\kappa] + \rho\ln\omega_{jc}\Big)\Big/\Big(1-\rho\Big)
\end{aligned} \tag{1.B.13}
$$

Equations (1.B.12) and (1.B.13) define the factor input demand conditional on local productivity and factor prices. We can now substitute the factor demand in the firm profit equation (1.B.10) and rewrite profits as a function of factor prices:

$$
\begin{aligned}
\Pi_{jc}^F &= Q^{1-\rho}\underbrace{\left(B_c\omega_{jc}N_{jc}^{\beta}M_{jc}^{1-\beta}\right)^{\rho}}_{=Y_{jc}} - w_c N_{jc} - r_c^M\tau_c\kappa M_{jc} \\
\Pi_{jc}^F(N_{jc}^*, M_{jc}^*) &= B_c^{\frac{\rho}{1-\rho}}\omega_{jc}^{\frac{\rho}{1-\rho}} w_c^{-\frac{\rho\beta}{1-\rho}} r_c^{M^{-\frac{\rho(1-\beta)}{1-\rho}}} (\tau_c\kappa)^{-\frac{\rho(1-\beta)}{1-\rho}} Q\rho^{\frac{\rho}{1-\rho}}\beta^{\frac{\rho\beta}{1-\rho}}(1-\beta)^{\frac{\rho(1-\beta)}{1-\rho}}(1-\rho)
\end{aligned}
$$

The term $1-\rho > 0$ at the end of the expression indicates that profits are a markup over costs. As defined before, this term is equivalent to the inverse of the absolute product demand elasticity, i.e., $1-\rho = -1/\epsilon^{\text{PD}}$. The more elastic product demand ($\epsilon^{\text{PD}} \downarrow$), the lower the markup and the lower firms' profits in the tradable good sector. Following Suárez Serrato and Zidar (2016) we define the value of firm j in city c in terms of factor costs and local productivity:

$$V_{jc}^F = \frac{1-\rho}{\rho} \ln \Pi_{jc}^F(N_{jc}^*, M_{jc}^*) = b_0 + \underbrace{\ln B_c - \beta \ln w_c - (1-\beta)\ln r_c^M - (1-\beta)\ln(\tau_c \kappa)}_{=V_c^F} + \ln \omega_{jc}$$

with constant $b_0 = \frac{1-\rho}{\rho} \ln Q + \ln \rho + \beta \ln \beta + (1-\beta)\ln(1-\beta) + \frac{1-\rho}{\rho} \ln(1-\rho)$. We assume that idiosyncratic productivity shifters $\ln \omega_{jc}$ are i.i.d. and follow an extreme value type I distribution with scale parameter σ^F. As before in the context of household location choice, we normalize the total number of firms to $F = 1$. Using the log-profit equation and the distributional assumption on $\ln \omega_{jc}$ we denote the share of firms locating in city c by:

$$F_c = \Pr\left(V_{jc}^F \geq V_{jk}^F, \forall k \neq c\right) = \frac{\exp\left(V_c^F/\sigma^F\right)}{\sum_{g=1}^C \exp\left(V_g^F/\sigma^F\right)}. \tag{1.B.14}$$

The number of firms in city c from equation (1.B.14) (extensive margin) and the firm-specific labor demand from equation (1.B.12) (intensive margin) define the aggregate log labor demand in city c:

$$\begin{aligned}
\ln N_c^D &= \ln F_c + \mathrm{E}_{\omega_{jc}}\left[\ln N_{jc}^*\right] \\
&= \frac{1}{\sigma^F}\ln B_c - \frac{\beta}{\sigma^F}\ln w_c - \frac{1-\beta}{\sigma^F}\ln r_c^M - \frac{1-\beta}{\sigma^F}\ln(\tau_c\kappa) - \ln\left(C\pi^F\right) \\
&\quad + \frac{\rho}{1-\rho}\ln B_c - \frac{1-\rho+\rho\beta}{1-\rho}\ln w_c - \frac{\rho(1-\beta)}{1-\rho}\ln r_c^M - \frac{\rho(1-\beta)}{1-\rho}\ln(\tau_c\kappa) \\
&\quad + \frac{1}{1-\rho}\ln\rho + \frac{1-\rho+\rho\beta}{1-\rho}\ln\beta + \frac{\rho(1-\beta)}{1-\rho}\ln(1-\beta) + \ln Q + \frac{\rho}{1-\rho}\mathrm{E}_{\omega_{jc}}\left[\ln\omega_{jc}\right] \\
\ln N_c^D &= \underbrace{\left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho}\right)}_{=\epsilon^{\text{B}}}\ln B_c \underbrace{-\left(1+\beta\left[\frac{1}{\sigma^F} + \frac{\rho}{1-\rho}\right]\right)}_{=\epsilon^{\text{ND}}}\ln w_c \underbrace{-(1-\beta)\left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho}\right)}_{=1+\epsilon^{\text{MD}}}\ln r_c^M \\
&\quad \underbrace{-(1-\beta)\left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho}\right)}_{=1+\epsilon^{\text{MD}}}\ln(\tau_c\kappa) + b_1
\end{aligned} \tag{1.B.15}$$

as a function of local productivity B_c, wages w_c and the (gross) factor price costs of commercial floor space $r_c^M \tau_c \kappa$ with constant term $b_1 = \Big(\ln\rho + [1-\rho+\rho\beta]\ln\beta + \rho[1-\beta]\ln[1-\beta] +$

$\rho \mathrm{E}_{\omega_{jc}} \left[\ln \omega_{jc}\right] \Big) \Big/ \Big(1-\rho\Big) + \ln Q - \ln C - \ln \pi^F$, where we define the average firm value across locations defined as $\pi^F = \frac{1}{C}\sum_{k=1}^{C} \exp\left(V_k^F/\sigma^F\right)$. The labor demand elasticity is defined as:

$$\frac{\ln N_c^D}{\ln w_c} = \underbrace{-\frac{\beta}{\sigma^F}}_{\text{Ext. margin}} \underbrace{-1-\frac{\beta\rho}{1-\rho}}_{\text{Int. margin}} = \epsilon^{\text{ND}} < 0. \tag{1.B.16}$$

Labor demand increases in local productivity B_c (i.e., $\epsilon^{\text{B}} > 0$) and decreases in the (gross) factor price of commercial floor space defined by $r_c^M \tau_c \kappa$ (i.e., $1 + \epsilon^{\text{MD}} < 0$).

Floor Space Demand. Analogous to labor demand, we can also derive firms' demand for commercial floor space using the intensive margin commercial floor space demand from equation (1.B.13) and the location choice of firms from equation (1.B.14):

$$\begin{aligned}
\ln M_c^D &= \ln F_c + \mathrm{E}_{\omega_{jc}}\left[\ln M_{jc}^*\right] \\
&= \frac{1}{\sigma^F}\ln B_c - \frac{\beta}{\sigma^F}\ln w_c - \frac{1-\beta}{\sigma^F}\ln r_c^M - \frac{1-\beta}{\sigma^F}\ln(\tau_c\kappa) - \ln\left(C\pi^F\right) \\
&\quad + \frac{\rho}{1-\rho}\ln B_c - \frac{\rho\beta}{1-\rho}\ln w_c - \frac{1-\rho\beta}{1-\rho}\ln r_c^M - \frac{1-\rho\beta}{1-\rho}\ln(\tau_c\kappa) \\
&\quad + \frac{1}{1-\rho}\ln\rho + \frac{\rho\beta}{1-\rho}\ln\beta + \frac{1-\rho\beta}{1-\rho}\ln(1-\beta) + \ln Q + \frac{\rho}{1-\rho}\mathrm{E}_{\omega_{jc}}\left[\ln\omega_{jc}\right] \\
\ln M_c^D &= \underbrace{\left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho}\right)}_{=\epsilon^{\text{B}}} \ln B_c \underbrace{-\beta\left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho}\right)}_{=1+\epsilon^{\text{ND}}} \ln w_c \underbrace{-\left(1 + [1-\beta]\left[\frac{1}{\sigma^F} + \frac{\rho}{1-\rho}\right]\right)}_{=\epsilon^{\text{MD}}} \ln r_c^M \\
&\quad \underbrace{-\left(1 + [1-\beta]\left[\frac{1}{\sigma^F} + \frac{\rho}{1-\rho}\right]\right)}_{=\epsilon^{\text{MD}}} \ln(\tau_c\kappa) + b_2
\end{aligned} \tag{1.B.17}$$

with constant $b_2 = \Big(\ln\rho + \rho\beta\ln\beta + [1-\rho\beta]\ln[1-\beta] + \rho\mathrm{E}_{\omega_{jc}}\left[\ln\omega_{jc}\right]\Big) \Big/ \Big(1-\rho\Big) + \ln Q - \ln C - \ln\pi^F$. The commercial floor space demand elasticity is defined as:

$$\frac{\partial \ln M_c^D}{\partial \ln r_c^M} = -\frac{1-\beta}{\sigma^F} - 1 - \frac{\rho(1-\beta)}{1-\rho} = \epsilon^{\text{MD}} < 0.$$

1.B.3 Construction Sector

We assume that a competitive, local construction sector provides both types of housing, residential and commercial floor space. Every municipality has positive supply of residential housing H_c and commercial floor space M_c. Following Ahlfeldt et al. (2015), we define the two types of floor space in terms of total floor space S_c available in city c:

$$H_c = \mu S_c \qquad\qquad M_c = (1-\mu)S_c \tag{1.B.18}$$

with residential share $\mu \in [0,1]$. This share is exogeneously given, and determined by the additional regulatory costs of commercial land, denoted $\phi \geq 1$. In equlibrium there must be a no-arbitrage condition between residential and commericial floor space:

$$r_c^M = \phi r_c^H \tag{1.B.19}$$

In line with Ahlfeldt et al. (2015), we assume that the observed floor price in the data is the maximum of residential and commercial rents, r_c^M. This implies that (i) observed residential rents are higher due existing regulatory costs, and (ii) both types of floor space are offered, $0 < \mu < 1$.

The construction sector relies on a Cobb-Douglas technology with constant returns to scale using land ready for construction L_c and capital K_c to produce total floor space $S_c = H_c + M_c$ (Epple et al., 2010). In contrast to the capital tax literature, we assume global capital markets with unlimited supply at an exogenous rate (Oates and Fischel, 2016). Consequently, the price for capital s is given and constant across municipalities. Profits in the construction industry are given by:

$$\Pi_c^C = r_c^M \underbrace{L_c^\gamma K_c^{1-\gamma}}_{=S_c} - l_c L_c - sK_c \tag{1.B.20}$$

with inputs and factor prices $L_c, K_c, l_c, s > 0$ and the output elasticity of land defined as $\gamma \in (0,1)$. Profit maximizing behavior yields the following first-order conditions:

$$\frac{\partial \Pi_c^C}{\partial L_c} = \gamma r_c^M \frac{S_c}{L_c} - l_c \stackrel{!}{=} 0$$
$$\frac{\partial \Pi_c^C}{\partial K_c} = (1-\gamma) r_c^M \frac{S_c}{K_c} - s \stackrel{!}{=} 0.$$

Treating the supply of capital K_c as infinitely elastic and the price of capital s as exogenous, we can solve for land prices l_c as a function of the floor space price r_c^M. Taking logs of the second first-order condition we can derive the capital demand of the construction industry conditional on factor prices and land input:

$$\ln s = \ln(1-\gamma) + \ln r_c^M + \ln S_c - \ln K_c$$

$$\begin{aligned}
\ln s &= \ln(1-\gamma) + \ln r_c^M + \gamma \ln L_c + (1-\gamma)\ln K_c - \ln K_c \\
\ln K_c &= \frac{1}{\gamma}\ln(1-\gamma) + \frac{1}{\gamma}\ln r_c^M + \ln L_c - \frac{1}{\gamma}\ln s.
\end{aligned}$$

Using the capital demand and the first-order condition with respect to land, we can solve for the price ratio of land to floor space in city c:

$$\begin{aligned}
\ln l_c &= \ln\gamma + \ln r_c^M + \ln S_c - \ln L_c \\
&= \ln\gamma + \ln r_c^M + \gamma\ln L_c + (1-\gamma)\ln K_c - \ln L_c \\
&= \ln\gamma + \ln r_c^M - (1-\gamma)\ln L_c + \frac{1-\gamma}{\gamma}\left(\ln(1-\gamma) + \ln r_c^M + \gamma\ln L_c - \ln s\right) \\
&= \underbrace{\ln\gamma + \frac{1-\gamma}{\gamma}\ln(1-\gamma)}_{=c_0} - \frac{1-\gamma}{\gamma}\ln s - \frac{1}{\gamma}\ln r_c^M \\
\ln l_c &= c_0 - \frac{1-\gamma}{\gamma}\ln s + \frac{1}{\gamma}\ln r_c^M
\end{aligned} \tag{1.B.21}$$

where we shorten notation by introducing the term c_0 that is constant across municipalities. Land prices increase in the floor space rent r_c^M (and equivalently in residential rents r_c^H).

1.B.4 Land Supply

While the total land area in each municipality is fixed and inelastic, the share of land ready for residential or commercial construction may be elastic. We model the supply of land ready for construction in city c according to the following log supply function:

$$\ln L_c = \theta \ln l_c \tag{1.B.22}$$

with land supply elasticity $\epsilon^{\text{LS}} = \theta > 0$. The preparation of new area includes, e.g., clearing and leveling the site, or building road access and connections to the electrical grid.

1.B.5 Local Governments

Local governments use share $\psi \in (0,1)$ of the property tax revenues to finance the local public good G_c. All remaining revenues are distributed lump-sum to all workers in the economy irrespective of location (share $1-\psi$). The government budget is defined as:

$$G_c = \psi \underbrace{\left(H_c r_c^H t_c + M_c r_c^M \left[\{1+t_c\}\kappa - 1\right]\right)}_{\text{Total tax revenue}}$$

$$\ln G_c = \ln \psi + \ln \left(H_c r_c^H t_c + M_c r_c^M \left[\{1 + t_c\} \kappa - 1 \right] \right), \tag{1.B.23}$$

where total tax revenue is the sum of residential property taxes, $H_c r_c^H t_c$, and property taxes on rented commercial floor space, $M_c r_c^M$. Increases in city c's property tax rate t_c yield higher tax revenues and thereby an mechanical increase in local spending on the public good.

1.B.6 Equilibrium

The spatial equilibrium is determined by equalizing supply and demand on the markets for labor, residential housing, commercial floor space and land in each city as well as the government budget constraint. Hence, we can summarize the equilibrium conditions using the following twelve equations:

$$\begin{aligned}
\ln N_c &= \frac{1-\delta}{\sigma^H} \ln w_c - \frac{\alpha(1-\delta)}{\sigma^H} \ln r_c^H - \frac{\alpha(1-\delta)}{\sigma^H} \ln \tau_c + \frac{1}{\sigma^H} \ln A_c + \frac{\delta}{\sigma^H} \ln G_c + a_1 \\
\ln N_c &= \left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right) \ln B_c - \left(1 + \beta \left[\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right] \right) \ln w_c - (1-\beta) \left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right) \ln r_c^M \\
&\quad - (1-\beta) \left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right) \ln(\tau_c \kappa) + b_1 \\
\ln H_c &= \ln N_c + \ln \alpha + \ln w_c - \ln r_c^H - \ln \tau_c \\
\ln H_c &= \ln \mu + \ln S_c \\
\ln M_c &= \left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right) \ln B_c - \beta \left(\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right) \ln w_c - \left(1 + [1-\beta] \left[\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right] \right) \ln r_c^M \\
&\quad - \left(1 + [1-\beta] \left[\frac{1}{\sigma^F} + \frac{\rho}{1-\rho} \right] \right) \ln(\tau_c \kappa) + b_2 \\
\ln M_c &= \ln(1-\mu) + \ln S_c \\
\ln S_c &= (1-\gamma) \ln K_c + \gamma \ln L_c \\
\ln K_c &= \ln L_c + \frac{1}{\gamma} \ln r_c^M + \frac{1}{\gamma} \ln(1-\gamma) - \frac{1}{\gamma} \ln s \\
\ln L_c &= \theta \ln l_c \\
\ln l_c &= c_0 - \frac{1-\gamma}{\gamma} \ln s + \frac{1}{\gamma} \ln r_c^M \\
\ln r_c^M &= \ln \phi + \ln r_c^H \\
\ln G_c &= \ln \psi + \ln \left(H_c r_c^H t_c + M_c r_c^M [\{1 + t_c\} \kappa - 1] \right)
\end{aligned}$$

where we again use $\tau_c = 1 + t_c$ to simplify the notation in the following. We further simplify the equations by using the key elasticities we defined above (see also Table 1.B.1 for an overview):

$$\ln N_c = \epsilon^{\text{NS}} \ln w_c + (1+\epsilon^{\text{HD}}) \ln r_c^H + (1+\epsilon^{\text{HD}}) \ln \tau_c + \epsilon^{\text{A}} \ln A_c + \delta \epsilon^{\text{A}} \ln G_c + a_1 \tag{1.B.24}$$

$$\ln N_c = \epsilon^{\text{B}} \ln B_c + \epsilon^{\text{ND}} \ln w_c + (1+\epsilon^{\text{MD}}) \ln r_c^M + (1+\epsilon^{\text{MD}}) \ln(\tau_c \kappa) + b_1 \tag{1.B.25}$$

$$\ln H_c = \ln N_c + \ln \alpha + \ln w_c - \ln r_c^H - \ln \tau_c \tag{1.B.26}$$

$$\ln H_c = \ln \mu + \ln S_c \tag{1.B.27}$$

$$\ln M_c = \epsilon^{\text{B}} \ln B_c + (1+\epsilon^{\text{ND}}) \ln w_c + \epsilon^{\text{MD}} \ln r_c^M + \epsilon^{\text{MD}} \ln(\tau_c \kappa) + b_2 \tag{1.B.28}$$

$$\ln M_c = \ln(1-\mu) + \ln S_c \tag{1.B.29}$$

$$\ln S_c = (1-\gamma) \ln K_c + \gamma \ln L_c \tag{1.B.30}$$

$$\ln K_c = \ln L_c + \frac{1}{\gamma} \ln r_c^M + \frac{1}{\gamma} \ln(1-\gamma) - \frac{1}{\gamma} \ln s \tag{1.B.31}$$

$$\ln L_c = \theta \ln l_c \tag{1.B.32}$$

$$\ln l_c = c_0 - \frac{1-\gamma}{\gamma} \ln s + \frac{1}{\gamma} \ln r_c^M \tag{1.B.33}$$

$$\ln r_c^M = \ln \phi + \ln r_c^H \tag{1.B.34}$$

$$\ln G_c = \ln \psi + \ln \left(H_c r_c^H t_c + M_c r_c^M [\{1 + t_c\}\kappa - 1] \right) \tag{1.B.35}$$

We can solve this system of equations for the equilibrium quantities in terms of population, residential housing, commercial floor space, use of capital, and developed land, equilibrium prices for labor, residential housing, commercial floor space, and land as well as public good provision in equilibrium.

Effective Housing Demand. To solve the model, we first derive the effective residential housing demand function, taking into account the extensive margin of people moving across locations. By combining equations (1.B.24) and (1.B.26), we get the following expression:

$$\ln H_c^D = a_1 + \ln \alpha + \epsilon^{\text{A}} \ln A_c + \delta \epsilon^{\text{A}} \ln G_c + \epsilon^{\text{HD}} \ln r_c^H + \epsilon^{\text{HD}} \ln \tau_c + \left(1 + \epsilon^{\text{NS}}\right) \ln w.$$

By clearing the labor market, i.e., equating expressions (1.B.24) and (1.B.25), we can derive wages as a function of amenities, public goods, and floor space prices:

$$\begin{aligned} \ln w_c = \Big(& b_1 - a_1 - \epsilon^{\text{A}} \ln A_c - \delta \epsilon^{\text{A}} \ln G_c + \epsilon^{\text{B}} \ln B_c - \left[1 + \epsilon^{\text{HD}}\right] \ln \tau_c + \left[1 + \epsilon^{\text{MD}}\right] \ln [\tau_c \kappa] \\ & - \left[1 + \epsilon^{\text{HD}}\right] \ln r_c^H + \left[1 + \epsilon^{\text{MD}}\right] \ln r_c^M \Big) \Big/ \left(\epsilon^{\text{NS}} - \epsilon^{\text{ND}}\right). \end{aligned} \tag{1.B.36}$$

Table 1.B.1: Key Elasticities of the Spatial Equilibrium Model

Key Elasticity	Definition
Panel A – Labor Market	
Labor Supply	
Wages	$\epsilon^{\text{NS}} = \frac{\partial \ln N_c}{\partial \ln w_c} = \frac{1-\delta}{\sigma^H}$
Exogenous Amenities	$\epsilon^{\text{A}} = \frac{\partial \ln N_c}{\partial \ln A_c} = \frac{1}{\sigma^H}$
Local Public Goods	$\delta\epsilon^{\text{A}} = \frac{\partial \ln N_c}{\partial \ln G_c} = \frac{\delta}{\sigma^H}$
Labor Demand	
Wages	$\epsilon^{\text{ND}} = \frac{\partial \ln N_c}{\partial \ln w_c} = -\left(1 + \beta\left[\frac{1}{\sigma^F} + \frac{\rho}{1-\rho}\right]\right)$
Productive Amenities	$\epsilon^{\text{B}} = \frac{\partial \ln N_c}{\partial \ln B_c} = \frac{1}{\sigma^F} + \frac{\rho}{1-\rho}$
Panel B – Construction Sector and Land Market	
Residential Housing Demand w.r.t. Rents	$\epsilon^{\text{HD}} = \frac{\partial \ln H_c}{\partial \ln r_c^H} = -\frac{\alpha(1-\delta)+\sigma^H}{\sigma^H}$
Commercial Floor Space Demand w.r.t. Rents	$\epsilon^{\text{MD}} = \frac{\partial \ln M_c}{\partial \ln r_c^M} = -\left(1 + [1-\beta]\left[\frac{1}{\sigma^F} + \frac{\rho}{1-\rho}\right]\right)$
Panel C – Land Market	
Land Supply w.r.t. Land Prices	$\frac{\partial \ln L_c}{\partial \ln l_c} = \theta$

Notes: This table summarizes the key supply and demand elasticities of the spatial equilibrium model.

As the partial derivative of log wages with respect to residential housing costs is positive ($-\left[1+\epsilon^{\text{HD}}\right]/\left[\epsilon^{\text{NS}}-\epsilon^{\text{ND}}\right] > 0$), wages (partly) compensate for higher rents and/or higher residential property taxes *ceteris paribus*. Using this intermediate wage equation, we can rewrite residential housing demand as a function of housing costs, exogenous amenities, and local public goods:

$$\begin{aligned}\ln H_c^D = \Big(&\left[1+\epsilon^{\text{NS}}\right] b_1 - \left[1+\epsilon^{\text{ND}}\right] a_1 - \epsilon^{\text{A}}\left[1+\epsilon^{\text{ND}}\right]\left[\ln A_c + \delta \ln G_c\right] + \epsilon^{\text{B}}\left[1+\epsilon^{\text{NS}}\right]\ln B_c \\ &- \left[1+\epsilon^{\text{NS}}+\epsilon^{\text{HD}}\left\{1+\epsilon^{\text{ND}}\right\}\right]\ln r_c^H - \left[1+\epsilon^{\text{NS}}+\epsilon^{\text{HD}}\left\{1+\epsilon^{\text{ND}}\right\}\right]\ln \tau_c \\ &+ \left[1+\epsilon^{\text{MD}}\right]\left[1+\epsilon^{\text{NS}}\right]\ln r_c^M + \left[1+\epsilon^{\text{MD}}\right]\left[1+\epsilon^{\text{NS}}\right]\ln[\tau_c\kappa]\Big)\Big/\Big(\epsilon^{\text{NS}}-\epsilon^{\text{ND}}\Big) + \ln\alpha\end{aligned}$$

and use the no-arbitrage condition in equation (1.B.34) to rewrite residential housing demand in terms of residential rents:

$$\begin{aligned}\ln H_c^D = \Big(&\left[1+\epsilon^{\text{NS}}\right] b_1 - \left[1+\epsilon^{\text{ND}}\right] a_1 + \left[1+\epsilon^{\text{MD}}\right]\left[1+\epsilon^{\text{NS}}\right]\ln\phi + \epsilon^{\text{B}}\left[1+\epsilon^{\text{NS}}\right]\ln B_c \\ &- \epsilon^{\text{A}}\left[1+\epsilon^{\text{ND}}\right]\Big[\ln A_c + \delta\ln G_c\Big] - \left[\epsilon^{\text{HD}}\left\{1+\epsilon^{\text{ND}}\right\} - \epsilon^{\text{MD}}\left\{1+\epsilon^{\text{NS}}\right\}\right]\ln r_c^H \\ &- \left[1+\epsilon^{\text{NS}}+\epsilon^{\text{HD}}\left\{1+\epsilon^{\text{ND}}\right\}\right]\ln\tau_c \\ &+ \left[1+\epsilon^{\text{MD}}\right]\left[1+\epsilon^{\text{NS}}\right]\ln[\tau_c\kappa]\Big)\Big/\Big(\epsilon^{\text{NS}}-\epsilon^{\text{ND}}\Big) + \ln\alpha.\end{aligned}$$

Residential housing demand is now a function of exogenous parameters and two endogenous measures, residential rents r_c^H and public good levels G_c.

Definition 1.B.1 (Effective Housing Demand). *The effective residential housing demand elasticity $\tilde{\epsilon}^{\mathrm{HD}}$ captures the response of residential housing demand to changes in residential rents holding public good levels constant but taking into account equilibrium effects on the labor market and the commercial floor space market. We define the effective residential housing demand elasticity as:*

$$\tilde{\epsilon}^{\mathrm{HD}} = -\frac{\epsilon^{\mathrm{HD}}[1+\epsilon^{\mathrm{ND}}] - \epsilon^{\mathrm{MD}}[1+\epsilon^{\mathrm{NS}}]}{\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}} < 0.$$

Given that $\epsilon^{\mathrm{HD}} < 0$, $\epsilon^{\mathrm{MD}} < 0$, $\epsilon^{\mathrm{ND}} < 0$, and $\epsilon^{\mathrm{NS}} > 0$, it follows that $\epsilon^{\mathrm{HD}} < 0$.

We can rewrite residential housing demand accordingly using this definition:

$$\begin{aligned} \ln H_c^D = & \Big(\left[1+\epsilon^{\mathrm{NS}}\right] b_1 - \left[1+\epsilon^{\mathrm{ND}}\right] a_1 + \left[1+\epsilon^{\mathrm{MD}}\right]\left[1+\epsilon^{\mathrm{NS}}\right] \ln \phi + \epsilon^{\mathrm{B}} \left[1+\epsilon^{\mathrm{NS}}\right] \ln B_c \\ & - \epsilon^{\mathrm{A}} \left[1+\epsilon^{\mathrm{ND}}\right] \left[\ln A_c + \delta \ln G_c\right] + \left[1+\epsilon^{\mathrm{MD}}\right]\left[1+\epsilon^{\mathrm{NS}}\right] \ln \kappa \Big) \Big/ \Big(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\Big) \\ & + \ln \alpha + \tilde{\epsilon}^{\mathrm{HD}} \ln r_c^H + \tilde{\epsilon}^{\mathrm{HD}} \ln \tau_c. \end{aligned} \tag{1.B.37}$$

Effective Housing Supply. To clear the residential housing market, demand needs to equal floor space supply, which we can rewrite as a function of capital costs and residential rents by combining equation (1.B.27) and equations (1.B.30)–(1.B.34):

$$\begin{aligned} \ln H_c^S &= \ln S_c + \ln \mu \\ &= \underbrace{(1-\gamma)\ln K_c + \gamma \ln L_c}_{=\ln S_c} + \ln \mu \\ &= \underbrace{(1-\gamma)\ln L_c + \frac{1-\gamma}{\gamma}\ln r_c^M + \frac{1-\gamma}{\gamma}\ln(1-\gamma) - \frac{1-\gamma}{\gamma}\ln s}_{=(1-\gamma)\ln K_c} + \gamma \ln L_c + \ln \mu \\ &= \underbrace{\theta \ln l_c}_{=\ln L_c} + \frac{1-\gamma}{\gamma}\ln r_c^M + \frac{1-\gamma}{\gamma}\ln(1-\gamma) - \frac{1-\gamma}{\gamma}\ln s + \ln \mu \\ &= \underbrace{\frac{\theta}{\gamma}\ln r_c^M - \frac{\theta(1-\gamma)}{\gamma}\ln s + \theta c_0}_{=\theta \ln l_c} + \frac{1-\gamma}{\gamma}\ln r_c^M + \frac{1-\gamma}{\gamma}\ln(1-\gamma) - \frac{1-\gamma}{\gamma}\ln s + \ln \mu \\ &= \frac{1-\gamma+\theta}{\gamma}\ln r_c^M - \frac{(1+\theta)(1-\gamma)}{\gamma}\ln s + \frac{1-\gamma}{\gamma}\ln(1-\gamma) + \theta c_0 + \ln \mu \end{aligned}$$

$$\ln H_c^S = \frac{1-\gamma+\theta}{\gamma} \underbrace{\left(\ln r_c^H + \ln \phi\right)}_{=\ln r_c^M} - \frac{(1+\theta)(1-\gamma)}{\gamma} \ln s + \frac{1-\gamma}{\gamma} \ln(1-\gamma) + \theta c_0 + \ln \mu.$$

Using these intermediate steps, we can also derive the effective housing supply elasticity.

Definition 1.B.2 (Effective Housing Supply). *The effective residential housing supply elasticity $\tilde{\epsilon}^{\mathrm{HS}}$ captures the response of residential housing supply to changes in residential rents taking into account both the factor substitution in the construction industry and the elasticity of land supply. We define the effective residential housing supply elasticity as:*

$$\tilde{\epsilon}^{\mathrm{HS}} = \frac{1-\gamma+\theta}{\gamma} > 0.$$

Given that $\gamma \in (0,1)$ and $\theta > 0$ it follows that $\tilde{\epsilon}^{\mathrm{HS}} > 0$.

By rewriting residential housing supply, we get:

$$\ln H_c^S = \tilde{\epsilon}^{\mathrm{HS}} \ln r_c^H + \tilde{\epsilon}^{\mathrm{HS}} \ln \phi - \frac{(1+\theta)(1-\gamma)}{\gamma} \ln s + \frac{1-\gamma}{\gamma} \ln(1-\gamma) + \theta c_0 + \ln \mu. \tag{1.B.38}$$

Rents. Using equations (1.B.37) and (1.B.38) we can clear the residential housing market and solve for equilibrium rents for residential floor space in city c as a function of equilibrium public good provision G_c^* and exogenous parameters:

$$\begin{aligned}
\ln r_c^{H*} = \Bigg(& \left[\ln \alpha - \ln \mu - \theta c_0 - \frac{1-\gamma}{\gamma} \ln\{1-\gamma\} + \frac{\{1+\theta\}\{1-\gamma\}}{\gamma} \ln s\right] \left[\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right] \\
& + \left[1+\epsilon^{\mathrm{NS}}\right] b_1 - \left[1+\epsilon^{\mathrm{ND}}\right] a_1 + \left[\left\{1+\epsilon^{\mathrm{MD}}\right\}\left\{1+\epsilon^{\mathrm{NS}}\right\} - \tilde{\epsilon}^{\mathrm{HS}}\left\{\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right\}\right] \ln \phi \\
& - \epsilon^{\mathrm{A}}\left[1+\epsilon^{\mathrm{ND}}\right]\left[\ln A_c + \delta \ln G_c^*\right] + \epsilon^{\mathrm{B}}\left[1+\epsilon^{\mathrm{NS}}\right] \ln B_c + \tilde{\epsilon}^{\mathrm{HD}}\left[\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right] \ln \tau_c \\
& + \left[1+\epsilon^{\mathrm{MD}}\right]\left[1+\epsilon^{\mathrm{NS}}\right] \ln \kappa \Bigg) \Bigg/ \Bigg(\left[\tilde{\epsilon}^{\mathrm{HS}} - \tilde{\epsilon}^{\mathrm{HD}}\right]\left[\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right] \Bigg)
\end{aligned}$$

$$\begin{aligned}
\ln r_c^{H*} = & \frac{\tilde{\epsilon}^{\mathrm{HD}}}{d_0} \ln \tau_c - \frac{\delta \epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{d_0\left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)} \ln G_c^* - \frac{\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{d_0\left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)} \ln A_c \\
& + \frac{\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)} \ln B_c + \frac{\left(1+\epsilon^{\mathrm{MD}}\right)\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)} \ln \kappa + \frac{d_{r^H}}{d_0}
\end{aligned} \tag{1.B.39}$$

with

$$d_0 = \tilde{\epsilon}^{\mathrm{HS}} - \tilde{\epsilon}^{\mathrm{HD}} > 0 \tag{1.B.40}$$

$$d_{r^H} = \ln \alpha - \ln \mu - \theta c_0 - \frac{1-\gamma}{\gamma} \ln(1-\gamma) + \frac{(1+\theta)(1-\gamma)}{\gamma} \ln s + \frac{1+\epsilon^{\mathrm{NS}}}{\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}} b_1$$

$$- \frac{1+\epsilon^{\mathrm{ND}}}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}} a_1 + \left(\frac{\left[1+\epsilon^{\mathrm{MD}}\right]\left[1+\epsilon^{\mathrm{NS}}\right]}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}} - \tilde{\epsilon}^{\mathrm{HS}} \right) \ln \phi.$$

Using the no-arbitrage condition in equation (1.B.34) we can solve for the equilibrium price of commercial floor space, again as a function of equilibrium local public goods:

$$\begin{aligned} \ln r_c^{M*} = & \frac{\tilde{\epsilon}^{\mathrm{HD}}}{d_0} \ln \tau_c - \frac{\delta \epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln G_c^* - \frac{\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln A_c \\ & + \frac{\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln B_c + \frac{\left(1+\epsilon^{\mathrm{MD}}\right)\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln \kappa + \frac{d_{r^M}}{d_0} \end{aligned} \tag{1.B.41}$$

with

$$\begin{aligned} d_{r^M} = & \ln \alpha - \ln \mu - \theta c_0 - \frac{1-\gamma}{\gamma} \ln(1-\gamma) + \frac{(1+\theta)(1-\gamma)}{\gamma} \ln s + \frac{1+\epsilon^{\mathrm{NS}}}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}} b_1 \\ & - \frac{1+\epsilon^{\mathrm{ND}}}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}} a_1 + \left(\frac{\left[1+\epsilon^{\mathrm{NS}}\right]\left[1+\epsilon^{\mathrm{MD}}\right]}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}} - \tilde{\epsilon}^{\mathrm{HD}} \right) \ln \phi. \end{aligned}$$

Wages. Having solved for the price of residential and commercial floor space, we can derive equilibrium wages in city c by exploiting the intermediate wage equation (1.B.36):

$$\begin{aligned} \ln w_c^* = & -\frac{\tilde{\epsilon}^{\mathrm{HS}}\left(\epsilon^{\mathrm{HD}}-\epsilon^{\mathrm{MD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln \tau_c - \frac{\delta \epsilon^{\mathrm{A}}\left(\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{MD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln G_c^* - \frac{\epsilon^{\mathrm{A}}\left(\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{MD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln A_c \\ & + \frac{\epsilon^{\mathrm{B}}\left(\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{HD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln B_c + \frac{\left(1+\epsilon^{\mathrm{MD}}\right)\left(\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{HD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln \kappa + \frac{d_w}{d_0} \end{aligned} \tag{1.B.42}$$

with

$$\begin{aligned} d_w = & \left(\theta c_0 - \ln \alpha + \frac{1-\gamma}{\gamma} \ln[1-\gamma] + \ln \mu - \frac{[1-\gamma][1+\theta]}{\gamma} \ln s \right) \frac{\epsilon^{\mathrm{HD}}-\epsilon^{\mathrm{MD}}}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}} \\ & + \frac{\tilde{\epsilon}^{\mathrm{HS}}\left(1+\epsilon^{\mathrm{HD}}\right)-\epsilon^{\mathrm{HD}}\left(1+\epsilon^{\mathrm{MD}}\right)}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}} \ln \phi - \frac{\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{MD}}}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}} a_1 + \frac{\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{HD}}}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}} b_1. \end{aligned}$$

Land Prices. The construction problem yields the relation between commercial floor space prices and land prices in equation (1.B.33). Solving for land prices yields:

$$\begin{aligned} \ln l_c^* = & \frac{\tilde{\epsilon}^{\mathrm{HD}}}{\gamma d_0} \ln \tau_c - \frac{\delta \epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln G_c^* - \frac{\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln A_c \\ & + \frac{\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln B_c + \frac{\left(1+\epsilon^{\mathrm{MD}}\right)\left(1+\epsilon^{\mathrm{NS}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} \ln \kappa + \frac{d_l}{\gamma d_0} \end{aligned} \tag{1.B.43}$$

with

$$\begin{aligned} d_l = {} & \ln\alpha - \ln\mu - \frac{1-\gamma}{\gamma}\ln(1-\gamma) - \frac{1+\epsilon^{\mathrm{ND}}}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}}a_1 + \frac{1+\epsilon^{\mathrm{NS}}}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}}b_1 + (\gamma d_0 - \theta)\,c_0 \\ & + \frac{(1-\gamma)(1+\theta-\gamma d_0)}{\gamma}\ln s + \frac{1+\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{HD}}\left(1+\epsilon^{\mathrm{ND}}\right)}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}}\ln\phi. \end{aligned}$$

Developed Land. Using equilibrium land prices and the land supply function allows to solve for equilibrium land use in city c:

$$\begin{aligned} \ln L_c^* = {} & \frac{\tilde{\epsilon}^{\mathrm{HD}}\theta}{\gamma d_0}\ln\tau_c - \frac{\delta\theta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\theta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\ & + \frac{\theta\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{\theta\left(1+\epsilon^{\mathrm{MD}}\right)\left(1+\epsilon^{\mathrm{NS}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln\kappa + \frac{\theta d_l}{\gamma d_0}. \end{aligned} \tag{1.B.44}$$

Capital Stock. Equilibrium land use and equilibrium floor space prices also determine the equilibrium capital stock in equation (1.B.31):

$$\begin{aligned} \ln K_c^* = {} & \frac{\tilde{\epsilon}^{\mathrm{HD}}(1+\theta)}{\gamma d_0}\ln\tau_c - \frac{\delta\epsilon^{\mathrm{A}}(1+\theta)\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\epsilon^{\mathrm{A}}(1+\theta)\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\ & + \frac{\epsilon^{\mathrm{B}}(1+\theta)\left(1+\epsilon^{\mathrm{NS}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{(1+\theta)\left(1+\epsilon^{\mathrm{MD}}\right)\left(1+\epsilon^{\mathrm{NS}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln\kappa + \frac{d_K}{\gamma d_0} \end{aligned} \tag{1.B.45}$$

with

$$\begin{aligned} d_K = {} & (1+\theta)\Big([\ln\alpha - \ln\mu]\left[\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right] - \left[1+\epsilon^{\mathrm{ND}}\right]a_1 + \left[1+\epsilon^{\mathrm{NS}}\right]b_1\Big) \\ & - \theta\,(1+\theta\gamma d_0)\,c_0 - \frac{(1-\gamma)(1+\theta)-\gamma d_0}{\gamma}\ln(1-\gamma) \\ & + \frac{(1-\gamma)(1+\theta)^2 - \gamma(1+\theta[1-\gamma])d_0}{\gamma}\ln s \\ & + \frac{(1+\theta)\left(1+\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{HD}}\left[1+\epsilon^{\mathrm{ND}}\right]\right)}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}}\ln\phi. \end{aligned}$$

Floor Space. Land use and the capital stock in equilibrium also determine total floor space production. Using the production function of the construction sector we can solve for the equilibrium floor space quantity in city c:

$$\begin{aligned} \ln S_c^* = {} & \frac{\tilde{\epsilon}^{\mathrm{HS}}\tilde{\epsilon}^{\mathrm{HD}}}{d_0}\ln\tau_c - \frac{\tilde{\epsilon}^{\mathrm{HS}}\delta\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\ & + \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{\tilde{\epsilon}^{\mathrm{HS}}\left(1+\epsilon^{\mathrm{MD}}\right)\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln\kappa + \frac{d_S}{\gamma d_0} \end{aligned} \tag{1.B.46}$$

with

$$
\begin{aligned}
d_S = \frac{1-\gamma+\theta}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}} \Big(& [\ln\alpha - \ln\mu]\left[\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right] - \left[1+\epsilon^{\mathrm{ND}}\right]a_1 + \left[1+\epsilon^{\mathrm{NS}}\right]b_1 \\
& + \left[1+\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{HD}}\left\{1+\epsilon^{\mathrm{ND}}\right\}\right]\ln\phi\Big) \\
& + (1-\gamma+\theta-\gamma d_0)\left(\frac{[1-\gamma][1+\theta]}{\gamma}\ln s - \theta c_0 - \frac{1-\gamma}{\gamma}\ln[1-\gamma]\right).
\end{aligned}
$$

Using the residential share μ of total floor space we can solve for residential housing in equilibrium:

$$
\begin{aligned}
\ln H_c^* = & \frac{\tilde{\epsilon}^{\mathrm{HS}}\tilde{\epsilon}^{\mathrm{HD}}}{d_0}\ln\tau_c - \frac{\tilde{\epsilon}^{\mathrm{HS}}\delta\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\
& + \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{\tilde{\epsilon}^{\mathrm{HS}}\left(1+\epsilon^{\mathrm{MD}}\right)\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln\kappa + \frac{d_H}{\gamma d_0}
\end{aligned}
\tag{1.B.47}
$$

with

$$d_H = d_S + \gamma d_0 \ln\mu.$$

Similarly we can solve for equilibrium commercial floor space production:

$$
\begin{aligned}
\ln M_c^* = & \frac{\tilde{\epsilon}^{\mathrm{HS}}\tilde{\epsilon}^{\mathrm{HD}}}{d_0}\ln\tau_c - \frac{\tilde{\epsilon}^{\mathrm{HS}}\delta\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\
& + \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{\tilde{\epsilon}^{\mathrm{HS}}\left(1+\epsilon^{\mathrm{MD}}\right)\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln\kappa + \frac{d_M}{\gamma d_0}
\end{aligned}
\tag{1.B.48}
$$

with

$$d_M = d_S + \gamma d_0 \ln(1-\mu).$$

Population. By exploiting the labor supply to city c as a function of rents and wages, we can also solve for equilibrium population:

$$
\begin{aligned}
\ln N_c^* = & -\frac{\tilde{\epsilon}^{\mathrm{HS}}\left(\epsilon^{\mathrm{ND}}\left[1+\epsilon^{\mathrm{HD}}\right]-\epsilon^{\mathrm{NS}}\left[1+\epsilon^{\mathrm{MD}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}\ln\tau_c - \frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{MD}}+\epsilon^{\mathrm{ND}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}\ln G_c^* \\
& - \frac{\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{MD}}+\epsilon^{\mathrm{ND}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}\ln A_c + \frac{\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{HD}}+\epsilon^{\mathrm{NS}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}\ln B_c \\
& + \frac{\left(1+\epsilon^{\mathrm{MD}}\right)\left(1+\epsilon^{\mathrm{HD}}+\epsilon^{\mathrm{NS}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}\ln\kappa + \frac{d_N}{d_0}
\end{aligned}
\tag{1.B.49}
$$

with

$$d_N = -\frac{1+\epsilon^{\mathrm{MD}}+\epsilon^{\mathrm{ND}}\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}}a_1 + \frac{1+\epsilon^{\mathrm{HD}}+\epsilon^{\mathrm{NS}}\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}}b_1$$

$$+\left(\ln\mu-\ln\alpha+\theta c_0+\frac{1-\gamma}{\gamma}\ln[1-\gamma]-\frac{[1-\gamma][1+\theta]}{\gamma}\ln s\right)$$
$$\times\left(\frac{\epsilon^{\text{ND}}\left[1+\epsilon^{\text{HD}}\right]-\epsilon^{\text{NS}}\left[1+\epsilon^{\text{MD}}\right]}{\epsilon^{\text{NS}}-\epsilon^{\text{ND}}}\right)$$
$$+\frac{\tilde{\epsilon}^{\text{HS}}\epsilon^{\text{ND}}\left(1+\epsilon^{\text{HD}}\right)+\left(1+\epsilon^{\text{MD}}\right)\left(1+\epsilon^{\text{HD}}+\epsilon^{\text{NS}}\right)}{\epsilon^{\text{NS}}-\epsilon^{\text{ND}}}\ln\phi.$$

Public Good Provision. So far, we solved the equilibrium conditional on equilibrium public good levels G_c^* in order to differentiate between the direct effects of taxes on equilibrium outcomes and the indirect effects operating through increases in local public goods financed via property taxes.

We can now also derive for equilibrium public good provision G_c^*. To simplify exposition and keep the model analytically tractable, we assume that rents for residential housing equal the prices for commercial floor space ($\phi = 1$), which implies that both types of land use are subject to the same regulations (Ahlfeldt et al., 2015). Moreover, we assume that residential and commercial floor space are taxed at the same rate, i.e., $\kappa = 1$.

Using the no-arbitrage condition from equation (1.B.34), the supply functions for residential and commercial floor space from equations (1.B.27) and (1.B.29), effective housing supply from equation (1.B.38), and equilibrium rents for residential housing in equation (1.B.51), we can solve for equilibrium public good provision:

$$\begin{aligned}
\ln G_c &= \ln\psi+\ln\left(H_c r_c^H t_c+M_c r_c^M\underbrace{\left[\{1+t_c\}^{\overbrace{\kappa}^{=1}}-1\right]}_{=t_c}\right)\\
&= \ln\psi+\ln\left(H_c r_c^H t_c+M_c\underbrace{\overbrace{\phi}^{=1}\; r_c^H}_{=r_c^M}\, t_c\right)\\
&= \ln\psi+\ln\left(\mu S_c r_c^H t_c+(1-\mu)S_c r_c^H t_c\right)\\
&= \ln\psi+\ln S_c+\ln r_c^H+\ln t_c\\
&= \ln\psi+\underbrace{\ln H_c-\ln\mu}_{=\ln S_c}+\ln r_c^H+\ln t_c
\end{aligned}$$

$$
\begin{aligned}
&= \underbrace{\tilde{\epsilon}^{\mathrm{HS}} \ln r_c^H + \tilde{\epsilon}^{\mathrm{HS}} \ln \phi - \frac{(1+\theta)(1-\gamma)}{\gamma} \ln s + \frac{1-\gamma}{\gamma} \ln(1-\gamma) + \theta c_0 + \ln \mu}_{=\ln H_c} \\
&\quad + \ln \psi - \ln \mu + \ln r_c^H + \ln t_c \\
&= \left(1 + \tilde{\epsilon}^{\mathrm{HS}}\right) \ln r_c^H + \ln t_c + \underbrace{\tilde{\epsilon}^{\mathrm{HS}} \ln \phi}_{=0} - \frac{(1+\theta)(1-\gamma)}{\gamma} \ln s + \frac{1-\gamma}{\gamma} \ln(1-\gamma) + \theta c_0 + \ln \psi \\
&= \left(1 + \tilde{\epsilon}^{\mathrm{HS}}\right) \left(\frac{\tilde{\epsilon}^{\mathrm{HD}}}{d_0} \ln \tau_c - \frac{\delta \epsilon^{\mathrm{A}} \left[1 + \epsilon^{\mathrm{ND}}\right]}{d_0 \left[\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right]} \ln G_c - \frac{\epsilon^{\mathrm{A}} \left[1 + \epsilon^{\mathrm{ND}}\right]}{d_0 \left[\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right]} \ln A_c \right. \\
&\qquad \left. + \frac{\epsilon^{\mathrm{B}} \left[1 + \epsilon^{\mathrm{NS}}\right]}{d_0 \left[\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right]} \ln B_c + \underbrace{\frac{\left[1 + \epsilon^{\mathrm{MD}}\right] \left[1 + \epsilon^{\mathrm{NS}}\right]}{d_0 \left[\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right]} \ln \kappa}_{=0} + \frac{d_{r^H}}{d_0} \right) \\
&\quad + \ln t_c - \frac{(1+\theta)(1-\gamma)}{\gamma} \ln s + \frac{1-\gamma}{\gamma} \ln(1-\gamma) + \theta c_0 + \ln \psi
\end{aligned}
$$

$$
\begin{aligned}
\ln G_c^* = &\left(\frac{\tilde{\epsilon}^{\mathrm{HD}} \left[1 + \tilde{\epsilon}^{\mathrm{HS}}\right]}{d_0} \ln \tau_c + \ln t_c - \frac{\epsilon^{\mathrm{A}} \left[1 + \tilde{\epsilon}^{\mathrm{HS}}\right] \left[1 + \epsilon^{\mathrm{ND}}\right]}{d_0 \left[\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right]} \ln A_c + \frac{d_r^H \left[1 + \tilde{\epsilon}^{\mathrm{HS}}\right]}{d_0} + d_G \right. \\
&\left. + \frac{\epsilon^{\mathrm{B}} \left[1 + \tilde{\epsilon}^{\mathrm{HS}}\right] \left[1 + \epsilon^{\mathrm{NS}}\right]}{d_0 \left[\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right]} \ln B_c \right) \Bigg/ \left(\frac{\delta \epsilon^{\mathrm{A}} \left[1 + \epsilon^{\mathrm{ND}}\right] \left[1 + \tilde{\epsilon}^{\mathrm{HS}}\right]}{d_0 \left[\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right]} + 1 \right) \qquad (1.B.50)
\end{aligned}
$$

with

$$
d_G = -\frac{(1+\theta)(1-\gamma)}{\gamma} \ln s + \frac{1-\gamma}{\gamma} \ln(1-\gamma) + \theta c_0 + \ln \psi.
$$

Summary. Hence, we arrive at the following spatial equilibrium prices and quantities for city c (conditional on equilibrium public good levels G_c^* and assuming equal tax rates on residential and commercial floor space, i.e., $\kappa = 1$):

$$
\begin{aligned}
\ln r_c^{H*} &= \frac{\tilde{\epsilon}^{\mathrm{HD}}}{d_0} \ln \tau_c - \frac{\delta \epsilon^{\mathrm{A}} \left(1 + \epsilon^{\mathrm{ND}}\right)}{d_0 \left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)} \ln G_c^* - \frac{\epsilon^{\mathrm{A}} \left(1 + \epsilon^{\mathrm{ND}}\right)}{d_0 \left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)} \ln A_c \\
&\quad + \frac{\epsilon^{\mathrm{B}} \left(1 + \epsilon^{\mathrm{NS}}\right)}{d_0 \left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)} \ln B_c + \frac{d_{r^H}}{d_0} \\
\ln r_c^{M*} &= \frac{\tilde{\epsilon}^{\mathrm{HD}}}{d_0} \ln \tau_c - \frac{\delta \epsilon^{\mathrm{A}} \left(1 + \epsilon^{\mathrm{ND}}\right)}{d_0 \left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)} \ln G_c^* - \frac{\epsilon^{\mathrm{A}} \left(1 + \epsilon^{\mathrm{ND}}\right)}{d_0 \left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)} \ln A_c \\
&\quad + \frac{\epsilon^{\mathrm{B}} \left(1 + \epsilon^{\mathrm{NS}}\right)}{d_0 \left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)} \ln B_c + \frac{d_{r^M}}{d_0}
\end{aligned}
$$

$$
\begin{aligned}
\ln l_c^* &= \frac{\tilde{\epsilon}^{\mathrm{HD}}}{\gamma d_0}\ln\tau_c - \frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\
&\quad + \frac{\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{d_l}{\gamma d_0} \\
\ln w_c^* &= -\frac{\tilde{\epsilon}^{\mathrm{HS}}\left(\epsilon^{\mathrm{HD}}-\epsilon^{\mathrm{MD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln\tau_c - \frac{\delta\epsilon^{\mathrm{A}}\left(\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{MD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\epsilon^{\mathrm{A}}\left(\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{MD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\
&\quad + \frac{\epsilon^{\mathrm{B}}\left(\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{HD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{d_w}{d_0} \\
\ln S_c^* &= \frac{\tilde{\epsilon}^{\mathrm{HS}}\tilde{\epsilon}^{\mathrm{HD}}}{d_0}\ln\tau_c - \frac{\tilde{\epsilon}^{\mathrm{HS}}\delta\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\
&\quad + \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{d_S}{\gamma d_0} \\
\ln H_c^* &= \frac{\tilde{\epsilon}^{\mathrm{HS}}\tilde{\epsilon}^{\mathrm{HD}}}{d_0}\ln\tau_c - \frac{\tilde{\epsilon}^{\mathrm{HS}}\delta\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\
&\quad + \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{d_H}{\gamma d_0} \\
\ln M_c^* &= \frac{\tilde{\epsilon}^{\mathrm{HS}}\tilde{\epsilon}^{\mathrm{HD}}}{d_0}\ln\tau_c - \frac{\tilde{\epsilon}^{\mathrm{HS}}\delta\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{A}}(1+\epsilon^{\mathrm{ND}})}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\
&\quad + \frac{\tilde{\epsilon}^{\mathrm{HS}}\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{d_M}{\gamma d_0} \\
\ln L_c^* &= \frac{\tilde{\epsilon}^{\mathrm{HD}}\theta}{\gamma d_0}\ln\tau_c - \frac{\delta\theta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\theta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\
&\quad + \frac{\theta\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{NS}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{\theta d_l}{\gamma d_0} \\
\ln N_c^* &= -\frac{\tilde{\epsilon}^{\mathrm{HS}}\left(\epsilon^{\mathrm{ND}}\left[1+\epsilon^{\mathrm{HD}}\right]-\epsilon^{\mathrm{NS}}\left[1+\epsilon^{\mathrm{MD}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}\ln\tau_c - \frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{MD}}+\epsilon^{\mathrm{ND}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}\ln G_c^* \\
&\quad - \frac{\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{MD}}+\epsilon^{\mathrm{ND}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}\ln A_c + \frac{\epsilon^{\mathrm{B}}\left(1+\epsilon^{\mathrm{HD}}+\epsilon^{\mathrm{NS}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{d_N}{d_0} \\
\ln K_c^* &= \frac{\tilde{\epsilon}^{\mathrm{HD}}(1+\theta)}{\gamma d_0}\ln\tau_c - \frac{\delta\epsilon^{\mathrm{A}}(1+\theta)\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* - \frac{\epsilon^{\mathrm{A}}(1+\theta)\left(1+\epsilon^{\mathrm{ND}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c \\
&\quad + \frac{\epsilon^{\mathrm{B}}(1+\theta)\left(1+\epsilon^{\mathrm{NS}}\right)}{\gamma d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c + \frac{d_K}{\gamma d_0} \\
\ln G_c^* &= \left(\frac{\tilde{\epsilon}^{\mathrm{HD}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]}{d_0}\ln\tau_c + \ln t_c - \frac{\epsilon^{\mathrm{A}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]\left[1+\epsilon^{\mathrm{ND}}\right]}{d_0\left[\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right]}\ln A_c + \frac{d_r^H\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]}{d_0} + d_G\right. \\
&\qquad \left. + \frac{\epsilon^{\mathrm{B}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]\left[1+\epsilon^{\mathrm{NS}}\right]}{d_0\left[\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right]}\ln B_c\right) \Bigg/ \left(\frac{\delta\epsilon^{\mathrm{A}}\left[1+\epsilon^{\mathrm{ND}}\right]\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]}{d_0\left[\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right]}+1\right)
\end{aligned}
$$

with d_0, d_{r^H}, d_{r^M}, d_l, d_w, d_S, d_H, d_M, d_N, and d_G being constant terms. From here, we can also derive the log real wage in city c using the equilibrium wage w_c^* and the equilibrium rent for residential housing r_c^{H*} (again conditional on equilibrium public good levels and assuming $\kappa = 1$):

$$\begin{aligned}\ln \frac{w_c^*}{r_c^{H*}\tau_c} = &-\frac{\tilde{\epsilon}^{\mathrm{HS}}\left(\epsilon^{\mathrm{HD}}-\epsilon^{\mathrm{MD}}+\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln \tau_c-\frac{\delta\epsilon^{\mathrm{A}}\left(\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{MD}}-\epsilon^{\mathrm{ND}}-1\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln G_c^* \\ &-\frac{\epsilon^{\mathrm{A}}\left(\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{MD}}-\epsilon^{\mathrm{ND}}-1\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln A_c+\frac{\epsilon^{\mathrm{B}}\left(\tilde{\epsilon}^{\mathrm{HS}}-\epsilon^{\mathrm{HD}}-\epsilon^{\mathrm{NS}}-1\right)}{d_0\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}\ln B_c\end{aligned} \tag{1.B.51}$$

1.B.7 Comparative Statics

Using the equilibrium outcomes derive above we can take a closer look at the comparative statics in the model. First we analyze the effects of tax increases on equilibrium prices and quantities. In a second step, we derive comparative statics with respect to the different amenities in the model.

Comparative Statics of Property Tax Increases. In the following, we derive how equilibrium outcomes respond to changes in property taxes. We derive the following theoretical predictions:

Lemma 1.B.1 (Public Goods). *The total effect of property tax increases on equilibrium public good provision in city c can be decomposed in (i) a direct, positive effect through higher revenues from taxing the existing housing stock at current prices, and (ii) an indirect, countervailing effect reflecting that higher taxes decrease prices and quantities traded on both floor space markets and thus the tax base.*

The higher the property tax rate t_c, the more important the second, indirect channel distorting the tax base relative to the direct revenue effect. The total effect of tax increases on public good spending will be positive as long as the tax rate is sufficiently small:

$$\frac{d\ln G_c^*}{d\ln \tau_c} > 0 \quad \Leftrightarrow \quad t_c < -\frac{\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}}{\tilde{\epsilon}^{\mathrm{HS}}\left(1+\tilde{\epsilon}^{\mathrm{HD}}\right)}.$$

Proof. The effect of property tax increases on equilibrium public good levels is given by:

$$\frac{d\ln G_c^*\left(t_c, r_c^{H*}\left[\tau_c, G_c^*\left\{\tau_c\right\}\right]\right)}{d\ln \tau_c} = \underbrace{\frac{\partial \ln G_c^*}{\partial \ln t_c}}_{>0}\underbrace{\frac{\partial \ln t_c}{\partial \ln \tau_c}}_{>0}+\underbrace{\frac{\partial \ln G_c^*}{\partial \ln r_c^{H*}}}_{>0}\left(\underbrace{\frac{\partial \ln r_c^{H*}}{\partial \ln \tau_c}}_{<0}+\underbrace{\frac{\partial \ln r_c^{H*}}{\partial \ln G_c^*}}_{<0}\frac{d\ln G_c^*}{d\ln \tau_c}\right)$$

$$\frac{d\ln G_c^*}{d\ln \tau_c} = \frac{\left(1+t_c\right)/t_c}{\frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}+1} + \frac{\tilde{\epsilon}^{\mathrm{HD}}\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right) \Big/ \left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)}{\frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}+1}$$

where the first fraction reflects the direct effect, the second fraction reflects the impact of property taxes on the housing market volume. While the numerator is positive for the direct and negative the indirect effect, respectively, the sign of the denominator and thus the sign of the total derivate depends on the model parameters. □

Lemma 1.B.2 (Rents)**.** *The total effect of property tax increases on equilibrium net rents for residential and commercial floor space in city c can be decomposed in (i) a direct, negative effect that is compensating for higher costs of living due to the tax increase, and (ii) an indirect effect operating through higher local public good provision that is theoretically undetermined. The indirect effect depends on the capitalization of public goods in rental prices and the degree to which tax increases raise the public good provision. It will be positive as long as public good spending increases in the tax rate.*

Proof. The effect of property tax increases on equilibrium rents is given by:

$$\begin{aligned}
\frac{d\ln r_c^{H*}\left(\tau_c, G_c^*\left[\tau_c\right]\right)}{d\ln \tau_c} &= \frac{d\ln r_c^{M*}\left(\tau_c, G_c^*\left[\tau_c\right]\right)}{d\ln \tau_c} \\
&= \frac{\partial \ln r_c^{H*}}{\partial \ln \tau_c} + \frac{\partial \ln r_c^{H*}}{\partial \ln G_c^*}\frac{d\ln G_c^*}{d\ln \tau_c} \\
&= \underbrace{\frac{\tilde{\epsilon}^{\mathrm{HD}}}{\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}}}_{<0} \underbrace{-\frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)}{\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}}_{>0} \underbrace{\frac{\frac{\tilde{\epsilon}^{\mathrm{HD}}\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}}+\frac{1+t_c}{t_c}}{\frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}+1}}_{\lesseqgtr 0},
\end{aligned}$$

where the first fraction reflects the direct, negative effect, the second fraction reflects the capitalization of public goods into rents, and the third fraction denotes the translation of property taxes into public good spending, which is theoretically undetermined. □

The statutory incidence of property taxes in our model is on the user of the housing services. Workers and firms thus have to finance the additional burden of higher property taxes. However, we assume that both groups of agents are at least somewhat mobile across jurisdictions and housing demand is thus at least somewhat elastic. As a result, renters are able to shift part of the additional tax burden onto landlords, which leads to a decrease in net rents for residential and commercial floor space when holding public good levels constant, i.e., a direct, negative effect. At the same time, tax increases impact the provision of local public goods in equilibrium.

Higher property taxes will increase tax revenues holding prices and quantities on the housing market fixed and thus increase the spending on public goods. Capitalization of public goods would thus reduce the downward pressure on net rents. However, there is a countervailing effect of property taxes on housing prices and quantities, which potentially lowers tax revenues and thereby public good spending. As discussed in Lemma 1.B.1, the combined effect of is theoretically undetermined, as is thus the indirect effect of property taxes on housing costs.

Lemma 1.B.3 (Wages). *The total effect of property tax increases on equilibrium wages in city c can be decomposed in (i) a direct effect that is potentially compensating for higher costs of living due to the tax increase, and (ii) an indirect effect operating through higher local public good provision that is theoretically undetermined. The direct effect depends on the relative strength of residential and commercial housing demand, i.e., the relative mobility of workers and firms to avoid higher taxes. The indirect effect depends on the capitalization of public goods in wages and the degree to which tax increases raise the public good provision. Both the direct and the indirect effect are theoretically undetermined.*

Proof. The effect of property tax increases on equilibrium wages is given by:

$$\begin{aligned}\frac{d\ln w_c^*\left(\tau_c, G_c^*\left[\tau_c\right]\right)}{d\ln\tau_c} &= \frac{\partial\ln w_c^*}{\partial\ln\tau_c} + \frac{\partial\ln w_c^*}{\partial\ln G_c^*}\frac{d\ln G_c^*}{d\ln\tau_c} \\ &= \underbrace{-\frac{\tilde{\epsilon}^{\mathrm{HS}}\left(\epsilon^{\mathrm{HD}} - \epsilon^{\mathrm{MD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)}}_{\lesseqgtr 0} \underbrace{-\frac{\delta\epsilon^{\mathrm{A}}\left(\tilde{\epsilon}^{\mathrm{HS}} - \epsilon^{\mathrm{MD}}\right)}{d_0\left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)}}_{<0} \underbrace{\frac{\frac{\tilde{\epsilon}^{\mathrm{HD}}\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}} + \frac{1+t_c}{t_c}}{\frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} + 1}}_{\lesseqgtr 0},\end{aligned}$$

where the first fraction reflects the direct effect, the second fraction reflects the capitalization of public goods into wages, and the third fraction denotes the translation of property taxes into public good spending. □

Tax increases trigger two opposing effects for profit maximizing firms in the city. On the one hand, higher property tax payments raise the factor price of commercial floor space and firms thus try to re-optimize by using less floor space relative to labor. On the other hand, property taxes make it more costly for workers to live in city c and residents demand higher wages to compensate for increased costs of living. Without compensating wage increases, inframarginal workers will move to other places. The sign and the magnitude of the two direct effects of tax increases on wages are determined by the relative strength of the residential and the commercial floor space demand elasticity, ϵ^{HD} and ϵ^{MD}, respectively. The indirect effect again operates through the capitalization of public goods into wages and depends on the extent to which tax increases yield additional public good spending at the local level.

Lemma 1.B.4 (Real Wages)**.** *The total effect of property tax increases on equilibrium real wages in city c, i.e., the wage adjusted for local costs of living, can be decomposed in (i) a direct, negative effect that reflects higher costs of living due to the tax increase even after accounting for potentially compensating rent decreases, and (ii) an indirect effect operating through higher local public good provision that is theoretically undetermined. The indirect effect depends on the capitalization of public goods in wages and rents, and the degree to which tax increases raise the public good provision, which is theoretically undetermined.*

Proof. The effect of property tax increases on equilibrium real wages is given by:

$$\frac{d\frac{w_c^*}{r_c^{H*}\tau_c}}{d\ln\tau_c} = \frac{\partial \ln w_c^*}{\partial \ln \tau_c} + \frac{\partial \ln w_c^*}{\partial \ln G_c^*}\frac{d\ln G_c^*}{d\ln\tau_c}$$

$$= \underbrace{-\frac{\tilde{\epsilon}^{\mathrm{HS}}\left(\epsilon^{\mathrm{HD}} - \epsilon^{\mathrm{MD}} + \epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)}{d_0\left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)}}_{<0} \underbrace{-\frac{\delta\epsilon^{\mathrm{A}}\left(\tilde{\epsilon}^{\mathrm{HS}} - \epsilon^{\mathrm{MD}} - \epsilon^{\mathrm{ND}} - 1\right)}{d_0\left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)}}_{<0} \underbrace{\frac{\frac{\tilde{\epsilon}^{\mathrm{HD}}\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}} + \frac{1+t_c}{t_c}}{\frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} + 1}}_{\lesseqgtr 0},$$

where the first fraction reflects the direct effect, the second fraction reflects the capitalization of public goods into wages and rents, and the third fraction denotes the translation of property taxes into public good spending. □

As seen before, net rents for residential housing may decrease in reaction to higher taxes thereby partly compensating for tax increases. The additional property tax burden would thus be shared between renters and landlords. Similarly, firms may compensate for higher costs of living in the municipality by paying higher wages. However, even taking together lower net rents and potentially higher wages does not fully balance the additional property tax burden. Real incomes in the jurisdiction thus decrease in response to tax increases (direct effect). For the case of real wages, the indirect effect operating through higher public good provision does not alleviate the direct effect, but yields additional downward pressure on real wages as long as the effect of property taxes on public good spending is positive. This mirrors the fact that workers compensation for higher costs of living may come through local public goods instead of higher real wages.

Lemma 1.B.5 (Population)**.** *The total effect of property tax increases on equilibrium population levels in city c can be decomposed in (i) a direct, negative effect that is due to lower real wages, and (ii) an indirect effect operating through higher local public good provision that is theoretically undetermined. The indirect effect depends on workers' (positive) valuation of public goods when choosing locations and the degree to which tax increases raise the public good provision. This* *indirect effect will be positive as long as public good spending increases in the tax rate.*

Proof. The effect of property tax increases on equilibrium population levels is given by:

$$\begin{aligned}
\frac{d \ln N_c^*\left(\tau_c, G_c^*\left[\tau_c\right]\right)}{d \ln \tau_c} &= \frac{\partial \ln N_c^*}{\partial \ln \tau_c} + \frac{\partial \ln N_c^*}{\partial \ln G_c^*}\frac{d \ln G_c^*}{d \ln \tau_c} \\
&= \underbrace{-\frac{\tilde{\epsilon}^{\mathrm{HS}}\left(\epsilon^{\mathrm{ND}}\left[1+\epsilon^{\mathrm{HD}}\right]-\epsilon^{\mathrm{NS}}\left[1+\epsilon^{\mathrm{MD}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}}_{<0} \\
&\quad \underbrace{-\frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{MD}}+\epsilon^{\mathrm{ND}}\left[1+\tilde{\epsilon}^{\mathrm{HS}}\right]\right)}{d_0\left(\epsilon^{\mathrm{NS}}+\epsilon^{\mathrm{ND}}\right)}}_{>0} \underbrace{\frac{\frac{\tilde{\epsilon}^{\mathrm{HD}}\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}}+\frac{1+t_c}{t_c}}{\frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}+1}}_{\lesseqgtr 0},
\end{aligned}$$

where the first fraction reflects the direct, negative effect, the second fraction reflects workers' valuation of public goods when choosing their location, and the third fraction denotes the translation of property taxes into public good spending, which is theoretically undetermined. □

When property taxes in city c increase, it becomes more expensive to live there – even after considering compensating effects though lower net rents and potentially higher wages. With constant local public goods and lower real incomes after the tax reform, the city becomes less attractive to live in (direct effect). As we assume that workers are at least somewhat mobile across jurisdictions, inframarginal workers will leave the municipality after the tax increase. The indirect effect through increases in local public goods works in the opposite direction and thus reduces the outflow of workers as long as public good levels increase in the tax rate.

Lemma 1.B.6 (Housing Stock)**.** *The total effect of property tax increases on the residential, commercial and total housing stock in equilibrium in city c can be decomposed in (i) a direct, negative effect that reflects lower rents and lower demand due to the tax increase, and (ii) an indirect effect operating through higher local public good provision that is theoretically undetermined. The indirect effect depends on the impact of public good supply on the local housing stock (positive) and the degree to which tax increases raise the public good provision. It will be positive as long as public good spending increases in the tax rate.*

Proof. The effect of property tax increases on the equilibrium housing stock is given by:

$$\begin{aligned}
\frac{d \ln H_c^*\left(\tau_c, G_c^*\left[\tau_c\right]\right)}{d \ln \tau_c} &= \frac{d \ln M_c^*\left(\tau_c, G_c^*\left[\tau_c\right]\right)}{d \ln \tau_c} = \frac{d \ln S_c^*\left(\tau_c, G_c^*\left[\tau_c\right]\right)}{d \ln \tau_c} \\
&= \frac{\partial \ln H_c^*}{\partial \ln \tau_c} + \frac{\partial \ln H_c^*}{\partial \ln G_c^*}\frac{d \ln G_c^*}{d \ln \tau_c}
\end{aligned}$$

$$= \underbrace{\frac{\tilde{\epsilon}^{\text{HS}}\tilde{\epsilon}^{\text{HD}}}{\tilde{\epsilon}^{\text{HS}} - \tilde{\epsilon}^{\text{HD}}}}_{<0} - \underbrace{\frac{\tilde{\epsilon}^{\text{HS}}\delta\epsilon^{\text{A}}\left(1+\epsilon^{\text{ND}}\right)}{\left(\tilde{\epsilon}^{\text{HS}} - \tilde{\epsilon}^{\text{HD}}\right)\left(\epsilon^{\text{NS}} - \epsilon^{\text{ND}}\right)}}_{>0} \underbrace{\frac{\frac{\tilde{\epsilon}^{\text{HD}}\left(1+\tilde{\epsilon}^{\text{HS}}\right)}{\tilde{\epsilon}^{\text{HS}}-\tilde{\epsilon}^{\text{HD}}} + \frac{1+t_c}{t_c}}{\frac{\delta\epsilon^{\text{A}}\left(1+\epsilon^{\text{ND}}\right)\left(1+\tilde{\epsilon}^{\text{HS}}\right)}{\left(\tilde{\epsilon}^{\text{HS}}-\tilde{\epsilon}^{\text{HD}}\right)\left(\epsilon^{\text{NS}}-\epsilon^{\text{ND}}\right)} + 1}}_{\lesseqgtr 0},$$

where the first fraction reflects the direct, negative effect, the second fraction reflects the impact of public goods on the housing stock, and the third fraction denotes the translation of property taxes into public good spending, which is theoretically undetermined. □

With constant public goods and lower real wages, the jurisdiction becomes less attractive to live in. Population levels decline in response to property tax increase. If less people are willing to locate in city c, the demand for residential housing declines. The same mechanism works for firms' location choice and their demand for commercial floor space. Eventually, both the residential housing stock and the amount of commercial floor space will be lower compared to the pre-reform equilibrium. This direct effect is in line with the prediction of the new view on the property tax. When accounting for endogenous local public goods, this prediction becomes less clear-cut due to the indirect effect. As long as public good spending increases in the tax rate, the public good provision alleviates the negative effect on the housing stock as higher public good levels increase the demand for city c despite the real wage loss.

Lemma 1.B.7 (Land Use)**.** *The total effect of property tax increases on equilibrium land use in city c can be decomposed in (i) a direct, negative effect that reflects lower activity in the construction sector due to the tax increase, and (ii) an indirect effect operating through higher local public good provision that is theoretically undetermined. The indirect effect depends on the impact of public goods on land use and the degree to which tax increases raise the public good provision. It will be positive as long as public good spending increases in the tax rate.*

Proof. The effect of property tax increases on equilibrium land use is given by:

$$\begin{aligned}
\frac{d\ln L_c^*\left(\tau_c, G_c^*\left[\tau_c\right]\right)}{d\ln\tau_c} &= \frac{\partial\ln L_c^*}{\partial\ln\tau_c} + \frac{\partial\ln L_c^*}{\partial\ln G_c^*}\frac{d\ln G_c^*}{d\ln\tau_c} \\
&= \underbrace{\frac{\tilde{\epsilon}^{\text{HD}}\theta}{\gamma\left(\tilde{\epsilon}^{\text{HS}} - \tilde{\epsilon}^{\text{HD}}\right)}}_{<0} - \underbrace{\frac{\theta\delta\epsilon^{\text{A}}\left(1+\epsilon^{\text{ND}}\right)}{\gamma\left(\tilde{\epsilon}^{\text{HS}} - \tilde{\epsilon}^{\text{HD}}\right)\left(\epsilon^{\text{NS}} - \epsilon^{\text{ND}}\right)}}_{>0} \underbrace{\frac{\frac{\tilde{\epsilon}^{\text{HD}}\left(1+\tilde{\epsilon}^{\text{HS}}\right)}{\tilde{\epsilon}^{\text{HS}}-\tilde{\epsilon}^{\text{HD}}} + \frac{1+t_c}{t_c}}{\frac{\delta\epsilon^{\text{A}}\left(1+\epsilon^{\text{ND}}\right)\left(1+\tilde{\epsilon}^{\text{HS}}\right)}{\left(\tilde{\epsilon}^{\text{HS}}-\tilde{\epsilon}^{\text{HD}}\right)\left(\epsilon^{\text{NS}}-\epsilon^{\text{ND}}\right)} + 1}}_{\lesseqgtr 0},
\end{aligned}$$

where the first fraction reflects the direct, negative effect, the second fraction reflects the impact of public goods on land use, and the third fraction denotes the translation of property taxes into public good spending, which is theoretically undetermined. □

With decreasing housing demand and lower levels of floor space production after the tax reform, the demand of the construction sector for land ready for construction decreases as well. This reflects the direct effect. As before, the indirect effect works in the opposite direction as long as public good spending increases in the tax rate.

Lemma 1.B.8 (Land Prices). *The total effect of property tax increases on equilibrium land prices in city c can be decomposed in (i) a direct, negative effect that reflects lower construction activity and lower land use in the construction sector due to the tax increase, and (ii) an indirect effect operating through higher local public good provision that is theoretically undetermined. The indirect effect depends on the impact of public goods on land prices and the degree to which tax increases raise the public good provision. It will be positive as long as public good spending increases in the tax rate.*

Proof. The effect of property tax increases on equilibrium land prices is given by:

$$\begin{aligned}\frac{d \ln l_c^* \left(\tau_c, G_c^* \left[\tau_c\right]\right)}{d \ln \tau_c} &= \frac{\partial \ln l_c^*}{\partial \ln \tau_c} + \frac{\partial \ln l_c^*}{\partial \ln G_c^*} \frac{d \ln G_c^*}{d \ln \tau_c} \\ &= \underbrace{\frac{\tilde{\epsilon}^{\mathrm{HD}}}{\gamma \left(\tilde{\epsilon}^{\mathrm{HS}} - \tilde{\epsilon}^{\mathrm{HD}}\right)}}_{<0} - \underbrace{\frac{\delta \epsilon^{\mathrm{A}} \left(1 + \epsilon^{\mathrm{ND}}\right)}{\gamma \left(\tilde{\epsilon}^{\mathrm{HS}} - \tilde{\epsilon}^{\mathrm{HD}}\right) \left(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\right)}}_{>0} \underbrace{\frac{\frac{\tilde{\epsilon}^{\mathrm{HD}}\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}} + \frac{1+t_c}{t_c}}{\frac{\delta \epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)} + 1}}_{\lesseqgtr 0},\end{aligned}$$

where the first fraction reflects the direct, negative effect, the second fraction reflects the impact of public goods on land prices, and the third fraction denotes the translation of property taxes into public good spending, which is theoretically undetermined. □

If population levels, floor space demand and the housing stock decrease, less land is needed for construction. As a result, land prices decrease as well to balance supply and demand, and to reach a new equilibrium on the market for land ready for development. This direct effect is again potentially diminished by an indirect effect operating through increases in local public goods.

Lemma 1.B.9 (Capital Stock). *The total effect of property tax increases on the equilibrium capital stock in city c can be decomposed in (i) a direct, negative effect that reflects lower construction activity due to the tax increase, and (ii) an indirect effect operating through higher local public good provision that is theoretically undetermined. The indirect effect depends on the impact of public goods on the capital stock and the degree to which tax increases raise the public good provision. It will be positive as long as public good spending increases in the tax rate.*

Proof. The effect of property tax increases on the equilibrium capital stock is given by:

$$\frac{d\ln K_c^*\left(\tau_c, G_c^*\left[\tau_c\right]\right)}{d\ln\tau_c} = \frac{\partial\ln K_c^*}{\partial\ln\tau_c} + \frac{\partial\ln K_c^*}{\partial\ln G_c^*}\frac{d\ln G_c^*}{d\ln\tau_c}$$
$$= \underbrace{\frac{\tilde{\epsilon}^{\mathrm{HD}}(1+\theta)}{\gamma\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)}}_{<0} - \underbrace{\frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)(1+\theta)}{\gamma\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}}_{>0}\underbrace{\frac{\frac{\tilde{\epsilon}^{\mathrm{HD}}\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}}+\frac{1+t_c}{t_c}}{\frac{\delta\epsilon^{\mathrm{A}}\left(1+\epsilon^{\mathrm{ND}}\right)\left(1+\tilde{\epsilon}^{\mathrm{HS}}\right)}{\left(\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}\right)\left(\epsilon^{\mathrm{NS}}-\epsilon^{\mathrm{ND}}\right)}+1}}_{\lesseqgtr 0},$$

where the first fraction reflects the direct, negative effect, the second fraction reflects the impact of public goods on the equilibrium capital stock, and the third fraction denotes the translation of property taxes into public good spending, which is theoretically undetermined. □

Lower population levels, lower housing demand and a smaller housing stock reduce the need for additional construction. Analogous to the demand for developed land, the capital demand of the construction sector declines as well. Again, this is in line with the capital tax view and reflects the direct effect holding public good provision constant. The indirect effect operates through the impact of public goods on the capital stock and will alleviate the direct negative effect as long as tax increases yield additional tax revenues that is spend on public goods.

1.B.8 Welfare Analysis

We assume a utilitarian welfare function that aggregates the utility of all agents in the economy:

$$W = W^H + W^F + \underbrace{W^C}_{=0} + W^L.$$

We measure worker welfare, W^H, by workers' utility and the welfare of firms owners, W^F, by the firm values defined above. The welfare of construction firm owners, W^C, and landlords' welfare, W^L, are measured by their profits. The construction sector is assumed to operate under perfect competition and makes zero profits, thus, welfare of construction firms is zero. We assume that the economy is large and a change in city c's property tax rate does not affect the utility of workers, firms or landlords in other locations. Following Kline and Moretti (2014) we define workers welfare as the inclusive value given by:

$$W^H = \sigma^H \ln\left(\sum_{c=1}^{C}\exp\left[\frac{V_c^H}{\sigma^H}\right]\right).$$

To first order, an increase in city c's property tax affects workers' welfare via the incidence of property taxes on gross rents $(1+t_c)r_c^{H*} = \tau_c r_c^{H*}$, its effect on wages w_c^*, and via the transmission into local public goods G_c^*:

$$\begin{aligned}
\frac{dW^H}{d\ln\tau_c} &= \frac{\sigma^H}{\sum_{k=1}^C \exp\left(V_k^H/\sigma^H\right)} \sum_{k=1}^C \frac{d\exp\left(V_k^H/\sigma^H\right)}{d\ln\tau_c} \\
&= \frac{\sigma^H}{\sum_{k=1}^C \exp\left(V_k^H/\sigma^H\right)} \sum_{k=1}^C \exp\left(\frac{V_k^H}{\sigma^H}\right) \frac{1}{\sigma^H} \frac{dV_k^H}{d\ln\tau_c} \\
&= \sum_{k=1}^C \frac{\exp\left(V_k^H/\sigma^H\right)}{\sum_{m=1}^C \exp\left(V_m^H/\sigma^H\right)} \frac{dV_k^H}{d\ln\tau_c} \\
&= \sum_{k=1}^C N_k \frac{dV_k^H}{d\ln\tau_c} \\
\frac{dW^H}{d\ln\tau_c} &= N_c \frac{dV_c^H}{d\ln\tau_c} = -N_c\left([1-\delta]\left[\alpha + \alpha\frac{d\ln r_c^{H*}}{d\ln\tau_c} - \frac{d\ln w_c^*}{d\ln\tau_c}\right] - \delta\frac{d\ln G_c^*}{d\ln\tau_c}\right).
\end{aligned}$$

The sign and the magnitude of this welfare effect for residents in city c depends (i) on the extent to which wages and net rents compensate for the real wage loss due to higher tax payments, and (ii) on the responsiveness of equilibrium public good spending to changes in the tax rate. The lower preferences for public goods, δ, the more important the former effect, the higher public good preferences, the more important the latter effect.

We derive firm values accordingly and again use the inclusive value to measure the welfare of firm owners (Suárez Serrato and Zidar, 2016):

$$W^F = \sigma^F \ln\left(\sum_{c=1}^C \exp\left[\frac{V_c^F}{\sigma^F}\right]\right).$$

A change in taxes in city c affects firm owners' welfare via the incidence on wages w_c^* and the impact on the gross price of commercial floor space $\kappa(1+t_c)r_c^{M*} = \kappa\tau_c r_c^{M*}$:

$$\frac{dW^F}{d\ln\tau_c} = F_c\frac{dV_c^F}{d\ln\tau_c} = -F_c\left([1-\beta] + [1-\beta]\frac{d\ln r_c^{M*}}{d\ln\tau_c} + \beta\frac{d\ln w_c^*}{d\ln\tau_c}\right).$$

The change in the welfare of firm owners depends on the share of the tax burden that can be passed on to landlords in terms of lower net prices for commercial floor space and the share that can be shifted to workers by lower wages.

The welfare of firm owners in the construction industry is given by their profits (see equation (1.B.20)):

$$W^C = \sum_{c=1}^{C} \Pi_c^C = \sum_{c=1}^{C} \left(r_c^{M*} S_c^* - sK_c^* - l_c^* L_c^* \right).$$

Property tax increases yield lower sales in the construction industry because workers and firms demand less floor space S_c^* and every unit is sold at a lower price r_c^{M*}. Construction firms react by decreasing their demand for land L_c^* and capital K_c^* and thus the price of land (l_c^*) will decrease as well. With some algebra, one can show that W^C evaluates to zero in equilibrium and construction firms still make zero profits irrespective of the tax.

We summarize the welfare of landlords by their profits as in Kline and Moretti (2014), i.e., the area between land prices and the inverse of the land supply function defined in equation (1.B.22), and scale it with the size of the nationwide land market denoted by Λ:

$$W^L = \frac{1}{\Lambda} \sum_{c=1}^{C} \int_0^{L_c^*} \left(l_c^* - u^{\frac{1}{\theta}} \right) \mathrm{d}u = \frac{1}{\Lambda} \sum_{c=1}^{C} \left(l_c^* L_c^* - \frac{L_c^{*\,1+\frac{1}{\theta}}}{1+\frac{1}{\theta}} \right) = \frac{1}{\Lambda} \sum_{c=1}^{C} \left(l_c^* L_c^* - \frac{\theta L_c^* \overbrace{L_c^{*\,\frac{1}{\theta}}}^{=l_c^*}}{1+\theta} \right)$$

$$W^L = \frac{1}{\Lambda} \sum_{c=1}^{C} \frac{l_c^* L_c^*}{1+\theta}.$$

Tax increases in city c reduce the welfare of landlords according to the following expression:

$$\begin{aligned}
\frac{dW^L}{d\ln\tau_c} &= \frac{1}{(1+\theta)\Lambda} \left(l_c^* \frac{dL_c^*}{d\ln\tau_c} + L_c^* \frac{dl_c^*}{d\ln\tau_c} \right) \\
&= \frac{1}{(1+\theta)\Lambda} \left(l_c^* \frac{d\exp\left[\ln L_c^*\right]}{d\ln\tau_c} + L_c^* \frac{d\exp\left[\ln l_c^*\right]}{d\ln\tau_c} \right) \\
&= \frac{1}{(1+\theta)\Lambda} \left(l_c^* \frac{d\exp\left[\ln L_c^*\right]}{d\ln L_c^*} \frac{d\ln L_c^*}{d\ln\tau_c} + L_c^* \frac{d\exp\left[\ln l_c^*\right]}{d\ln l_c^*} \frac{d\ln l_c^*}{d\ln\tau_c} \right) \\
&= \frac{1}{(1+\theta)\Lambda} \left(l_c^* L_c^* \frac{d\ln L_c^*}{d\ln\tau_c} + L_c^* l_c^* \frac{d\ln l_c^*}{d\ln\tau_c} \right) \\
&= \frac{l_c^* L_c^*}{(1+\theta)\Lambda} \left(\frac{d\ln L_c^*}{d\ln\tau_c} + \frac{d\ln l_c^*}{d\ln\tau_c} \right) \\
\frac{dW^L}{d\ln\tau_c} &= \frac{l_c^* L_c^*}{\gamma\Lambda} \frac{d\ln r_c^H}{d\ln\tau_c} = \frac{l_c^* L_c^*}{\Lambda} \frac{d\ln l_c^*}{d\ln\tau_c} = \Lambda_c \frac{d\ln l_c^*}{d\ln\tau_c},
\end{aligned}$$

where Λ_c denotes the share of local land sales $l_c^* L_c^*$ relative to the nationwide land market Λ.

The stronger the incidence of property taxes on land prices and the more severe the reduction in land demand due to higher taxes, the bigger the welfare loss for landlords. As landlords' welfare is decreasing in the land supply elasticity (see denominator), landlords will only bear part of the tax burden as long as the supply of land ready for construction is not perfectly elastic. Otherwise landlords make zero profits and won't bear any tax burden.

1.B.9 The Property Tax as a Specific Tax

So far we assumed that property is taxed *ad valorem*, which allows us to derive an analytical solution for the spatial equilibrium. However, our central incidence prediction of a direct, negative effect of property taxes on net rents compensating for higher costs of living does not rely on this assumption but also goes through when we model property taxes as a specific tax. To see this, consider an alternative formulation of the households' budget constraint in equation (1.B.1):

$$(r_c^H + t_c)h_i + px_i = w_c$$

and an alternative profit specification for firms in the tradable good sector (see equation (1.B.9)):

$$\Pi_{jc}^F = p_{jc}Y_{jc} - w_c N_{jc} - (r_c^M + t_c\kappa)M_{jc}.$$

Using these assumptions, we can derive alternative functions for labor supply, residential housing demand, labor demand, and commercial floor space demand using the key elasticities defined above:

$$\ln N_c^S = \epsilon^{\mathrm{NS}} \ln w_c + \left(1 + \epsilon^{\mathrm{HD}}\right) \ln\left(r_c^H + t_c\right) + \epsilon^{\mathrm{A}} \ln A_c + \delta\epsilon^{\mathrm{A}} \ln G_c + a_1 \tag{1.B.52}$$

$$\ln H_c^D = \ln N_c + \ln\alpha + \ln w_c - \ln\left(r_c^H + t_c\right) \tag{1.B.53}$$

$$\ln N_c^D = \epsilon^{\mathrm{B}} \ln B_c + \epsilon^{\mathrm{ND}} \ln w_c + \left(1 + \epsilon^{\mathrm{MD}}\right) \ln\left(r_c^M + t_c\kappa\right) + b_1 \tag{1.B.54}$$

$$\ln M_c^D = \epsilon^{\mathrm{B}} \ln B_c + \left(1 + \epsilon^{\mathrm{ND}}\right) \ln w_c + \epsilon^{\mathrm{MD}} \ln\left(r_c^M + t_c\kappa\right) + b_2. \tag{1.B.55}$$

The equilibrium is now characterized by equations (1.B.27), (1.B.29)-(1.B.35) and (1.B.52)-(1.B.55). As in the case of an *ad valorem* tax, we also derive effective residential housing demand and effective residential housing supply as functions of residential rents and property taxes when modeling the property tax as a specific tax. Using the alternative expressions for labor demand and labor supply, we can derive new equilibrium wages conditional on taxes, floor space prices, and amenities:

$$\ln w_c = \Big(b_1 - a_1 - \epsilon^{\mathrm{A}} \ln A_c - \delta\epsilon^{\mathrm{A}} \ln G_c + \epsilon^{\mathrm{B}} \ln B_c - \left[1 + \epsilon^{\mathrm{HD}}\right] \ln \left[r_c^H + t_c\right] \\ + \left[1 + \epsilon^{\mathrm{MD}}\right] \ln \left[r_c^M + t_c \kappa\right] \Big) \Big/ \Big(\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}\Big).$$

Using this intermediate step and the equilibrium conditions from above, we can solve for residential housing demand as a function of residential housing costs r_c^H, taxes, and amenities (note that we set $\phi = 1$ and $\kappa = 1$ again to keep the model analytically tractable):

$$\begin{aligned}
\ln H_c^D &= \ln N_c + \ln \alpha + \ln w_c - \ln \left(r_c^H + t_c\right) \\
&= \overbrace{\epsilon^{\mathrm{NS}} \ln w_c + \left(1 + \epsilon^{\mathrm{HD}}\right) \ln \left(r_c^H + t_c\right) + \epsilon^{\mathrm{A}} \ln A_c + \delta\epsilon^{\mathrm{A}} \ln G_c + a_1}^{=\ln N_c^D} \\
&\quad + \ln \alpha + \ln w_c - \ln \left(r_c^H + t_c\right) \\
&= \left(1 + \epsilon^{\mathrm{NS}}\right) \ln w_c + \epsilon^{\mathrm{HD}} \ln \left(r_c^H + t_c\right) + \epsilon^{\mathrm{A}} \ln A_c + \delta\epsilon^{\mathrm{A}} \ln G_c + \ln \alpha + a_1
\end{aligned}$$

$$\begin{aligned}
\ln H_c^D = \ln \alpha &+ \frac{1 + \epsilon^{\mathrm{NS}}}{\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}} b_1 - \frac{1 + \epsilon^{\mathrm{ND}}}{\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}} a_1 - \overbrace{\frac{\epsilon^{\mathrm{HD}}\left(1 + \epsilon^{\mathrm{ND}}\right) - \epsilon^{\mathrm{MD}}\left(1 + \epsilon^{\mathrm{NS}}\right)}{\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}}}^{=\tilde{\epsilon}^{\mathrm{HD}}} \ln \left(r_c^H + t_c\right) \\
&- \frac{1 + \epsilon^{\mathrm{ND}}}{\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}} \epsilon^{\mathrm{A}} \ln A_c - \frac{1 + \epsilon^{\mathrm{ND}}}{\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}} \delta\epsilon^{\mathrm{A}} \ln G_c + \frac{1 + \epsilon^{\mathrm{NS}}}{\epsilon^{\mathrm{NS}} - \epsilon^{\mathrm{ND}}} \epsilon^{\mathrm{B}} \ln B_c.
\end{aligned}$$

We can also write residential housing supply as a function of residential rents and exogenous model parameters only. It is still given by:

$$\ln H_c^S = \tilde{\epsilon}^{\mathrm{HS}} \ln r_c^H + \theta c_0 + \frac{1 - \gamma}{\gamma} \ln(1 - \gamma) + \ln \mu - \frac{(1 - \gamma)(1 + \theta)}{\gamma} \ln s.$$

In the spatial equilibrium, demand and supply of residential housing must be equal. Using the above equations we can solve for equilibrium rents for residential floor space. Now consider the introduction of a property tax in city c and derive the tax incidence $\frac{dr_c^H}{dt_c}$ (ignoring the supply of public goods):

$$\begin{aligned}
H_c^S\left(r_c^H\right) &= H_c^D\left(r_c^H + t_c\right) \\
\frac{dH_c^S\left(r_c^H\right)}{dt_c} &= \frac{dH_c^D\left(r_c^H + t_c\right)}{dt_c} \\
\frac{dH_c^S\left(r_c^H\right)}{dr_c^H} \frac{dr_c^H}{dt_c} &= \frac{dH_c^D\left(r_c^H + t_c\right)}{dr_c^H} \left(1 + \frac{dr_c^H}{dt_c}\right)
\end{aligned}$$

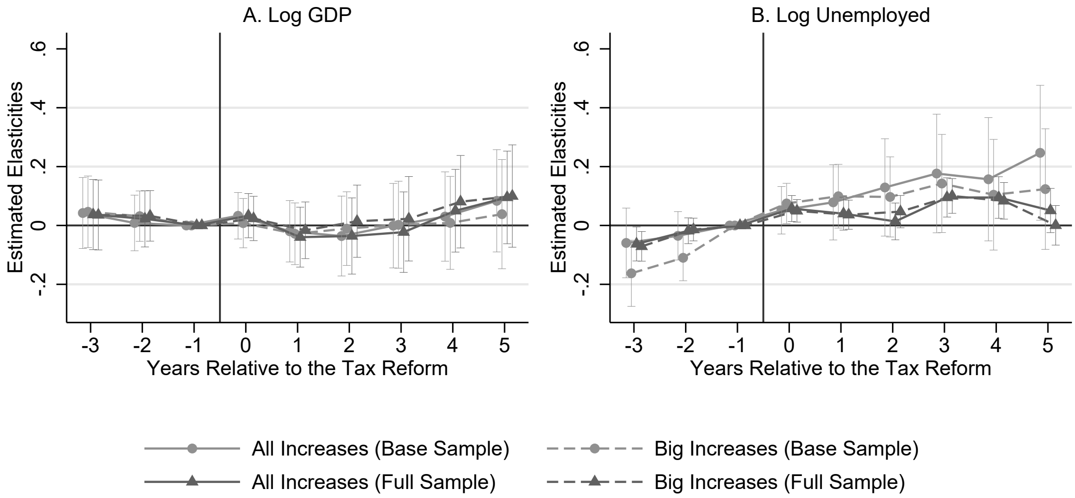
$$
\begin{aligned}
\frac{dr_c^H}{dt_c}\left(\frac{dH_c^S\left[r_c^H\right]}{dr_c^H}-\frac{dH_c^D\left[r_c^H+t_c\right]}{dr_c^H}\right)=\frac{dH_c^D\left(r_c^H+t_c\right)}{dr_c^H}\\
\frac{dr_c^H}{dt_c}=\frac{\frac{dH_c^D\left(r_c^H+t_c\right)}{dr_c^H}}{\frac{dH_c^S\left(r_c^H\right)}{dr_c^H}-\frac{dH_c^D\left(r_c^H+t_c\right)}{dr_c^H}}\\
=\frac{\frac{dH_c^D}{dr_c^H}\frac{r_c^H}{H_c^D}}{\frac{dH_c^S}{dr_c^H}\frac{r_c^H}{H_c^S}-\frac{dH_c^D}{dr_c^H}\frac{r_c^H}{H_c^D}}\\
\frac{dr_c^H}{dt_c}=\frac{\tilde{\epsilon}^{\mathrm{HD}}}{\tilde{\epsilon}^{\mathrm{HS}}-\tilde{\epsilon}^{\mathrm{HD}}},
\end{aligned}
\tag{1.B.56}
$$

which corresponds to the direct effect in the case of an *ad valorem* tax in Lemma 1.B.2. In both cases – for specific taxes as well as *ad valorem* taxes – the incidence after introducing a new tax depends on the relative size of the effective housing supply and the effective housing demand elasticity with respect to the rental price.

Appendix 1.C Additional Results

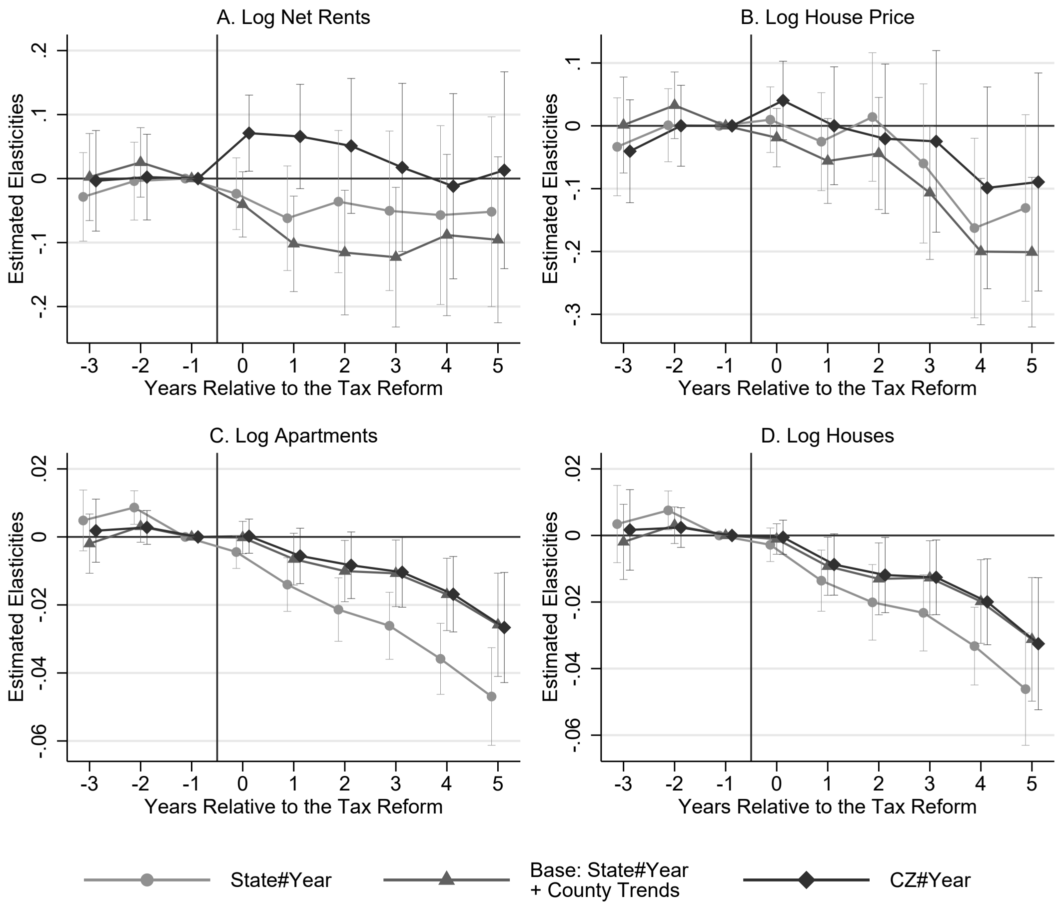
1.C.1 Robustness Checks

Figure 1.C.1: The Effects of Property Taxes on the Business Cycle



Notes: This figure shows the effects of property taxes on the business cycle using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases), or as an increase that is greater than or equal to the median of the tax change distribution (big increases). The base sample includes all municipalities from our housing data set (see Section 1.4.2 for details), the full sample includes all municipalities for which we have data on the respective outcome. Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.2: The Effects of Property Taxes on the Housing Market By Regional Controls

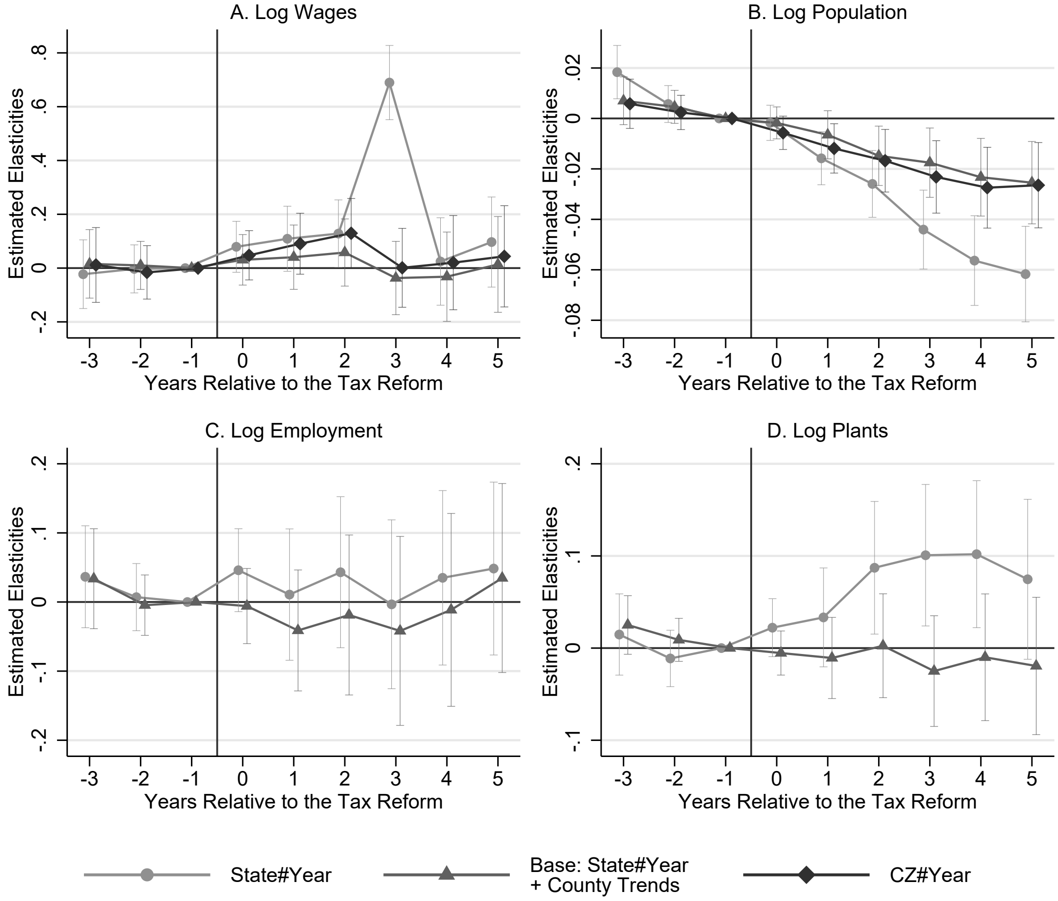


Notes: This figure shows the effects of property taxes on the housing market by regional controls using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.3: The Effects of Property Taxes on the Land Market By Regional Controls

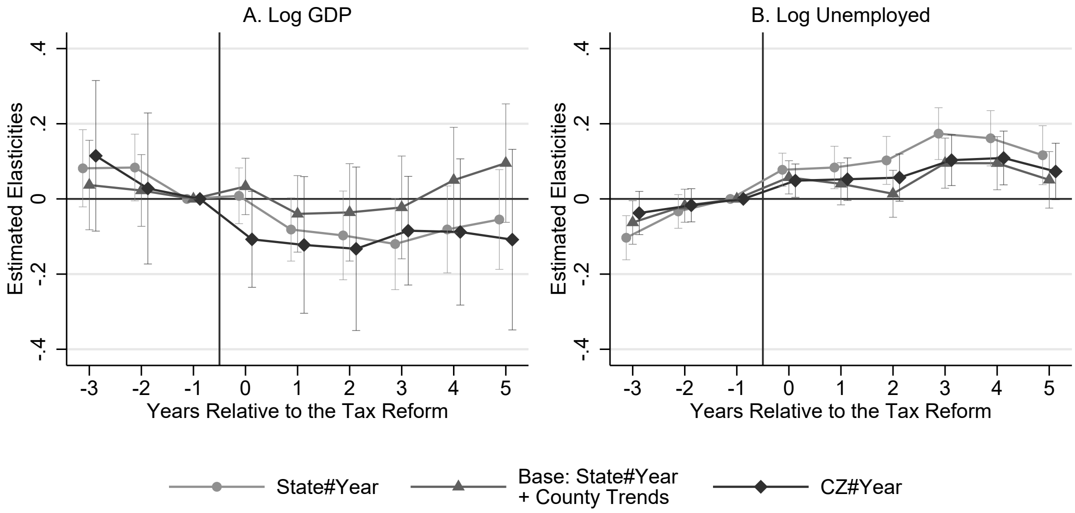
Notes: This figure shows the effects of property taxes on the land market by regional controls using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.4: The Effects of Property Taxes on the Labor Market By Regional Controls



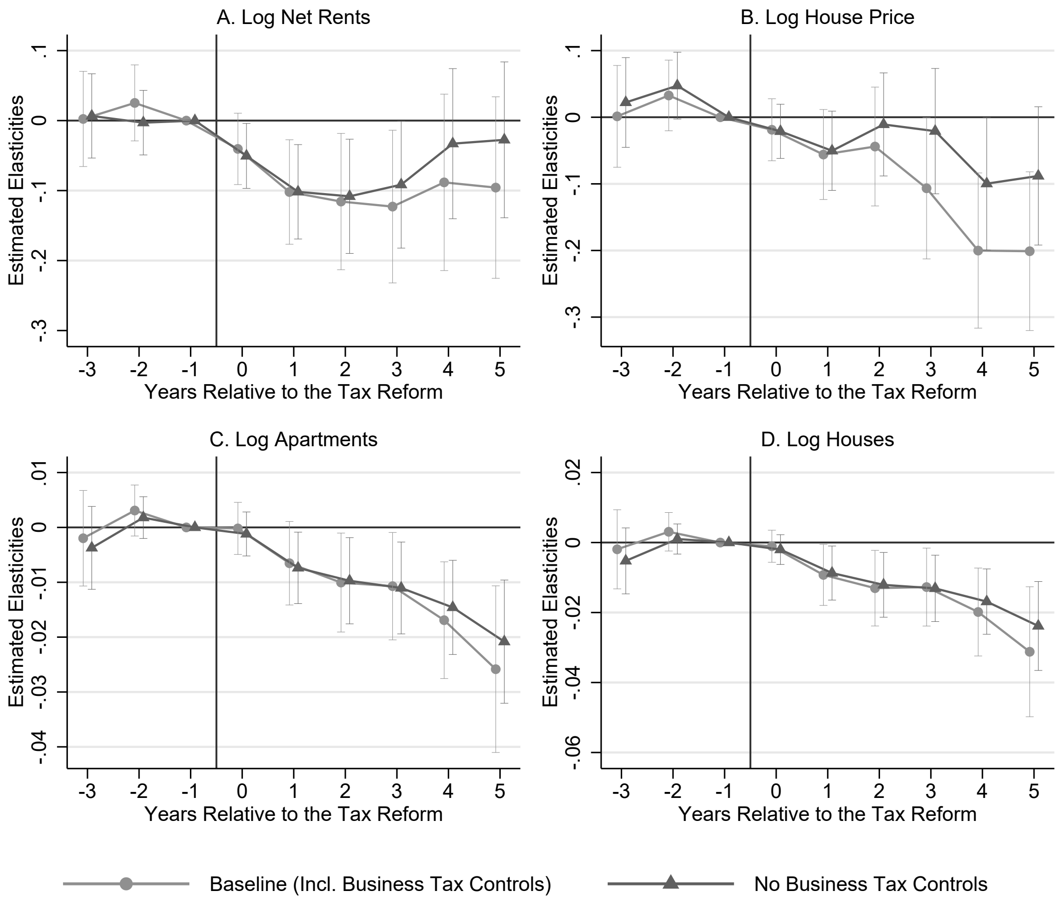
Notes: This figure shows the effects of property taxes on the labor market by regional controls using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Panels C and D are based on a subsample of large municipalities ("urban counties") as data on employment and plants are only observed at the county level. Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.5: The Effects of Property Taxes on the Business Cycle By Regional Controls



Notes: This figure shows the effects of property taxes on the business cycle by regional controls using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.6: The Effects of Property Taxes on the Housing Market By Business Tax Controls

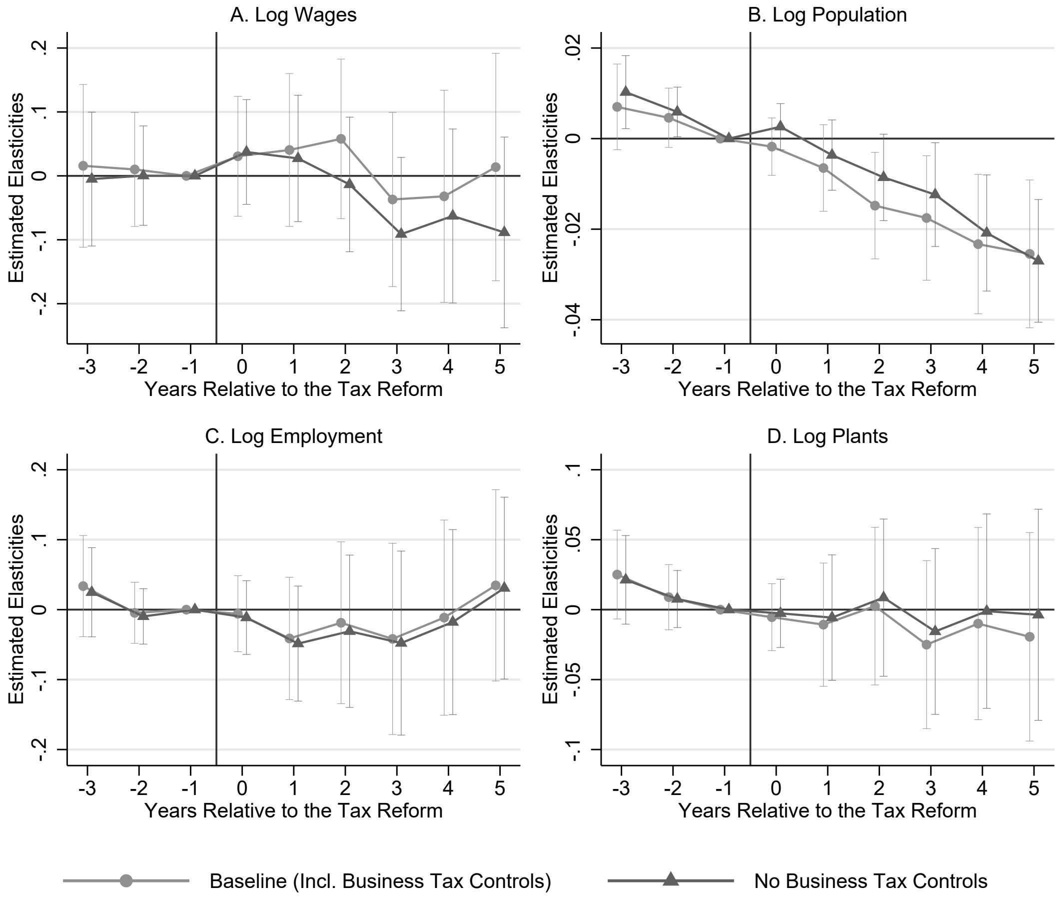


Notes: This figure shows the effects of property taxes on the housing market by business tax controls using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.7: The Effects of Property Taxes on the Land Market By Business Tax Controls

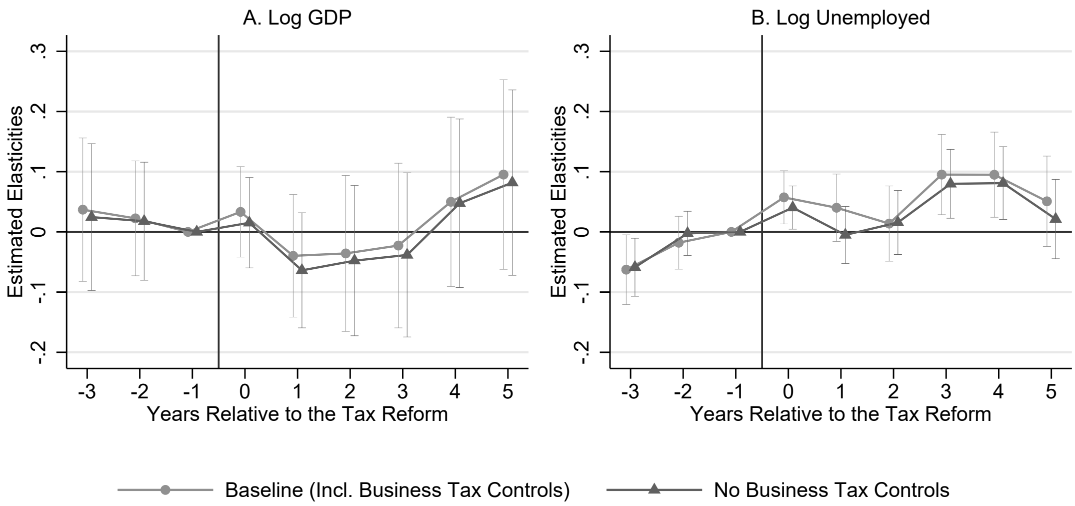
Notes: This figure shows the effects of property taxes on the land market by business tax controls using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.8: The Effects of Property Taxes on the Labor Market By Business Tax Controls



Notes: This figure shows the effects of property taxes on the labor market by business tax controls using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Panels C and D are based on a subsample of large municipalities ("urban counties") as data on employment and plants are only observed at the county level. Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

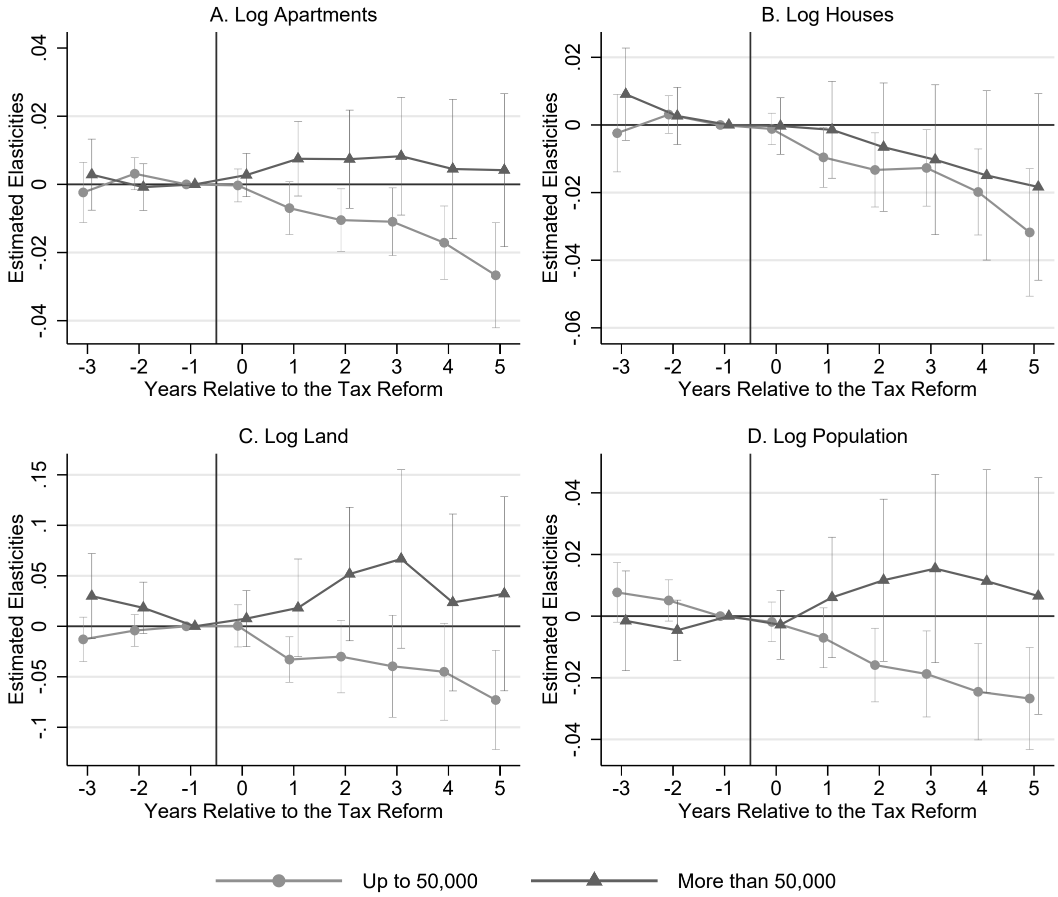
Figure 1.C.9: The Effects of Property Taxes on the Business Cycle By Business Tax Controls



Notes: This figure shows the effects of property taxes on the business cycle by business tax controls using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

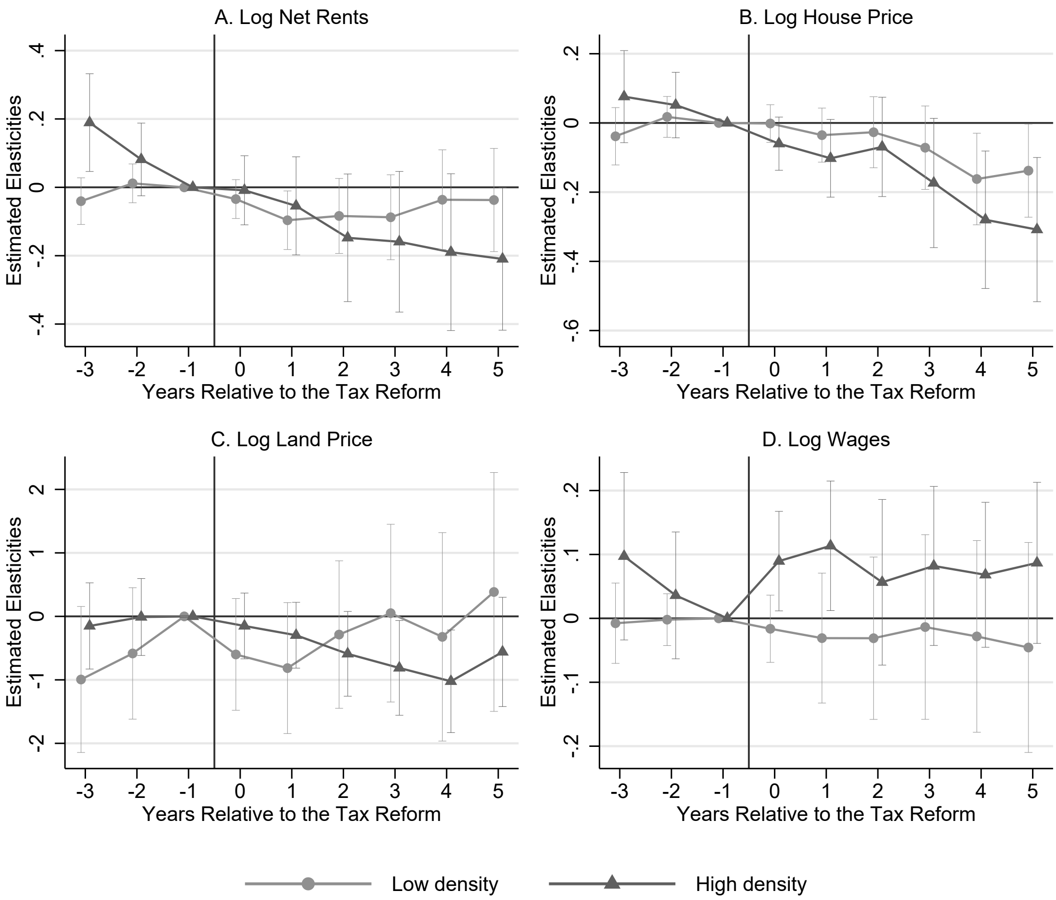
1.C.2 Heterogeneous Effects

Figure 1.C.10: The Effects of Property Taxes on Quantities by City Size



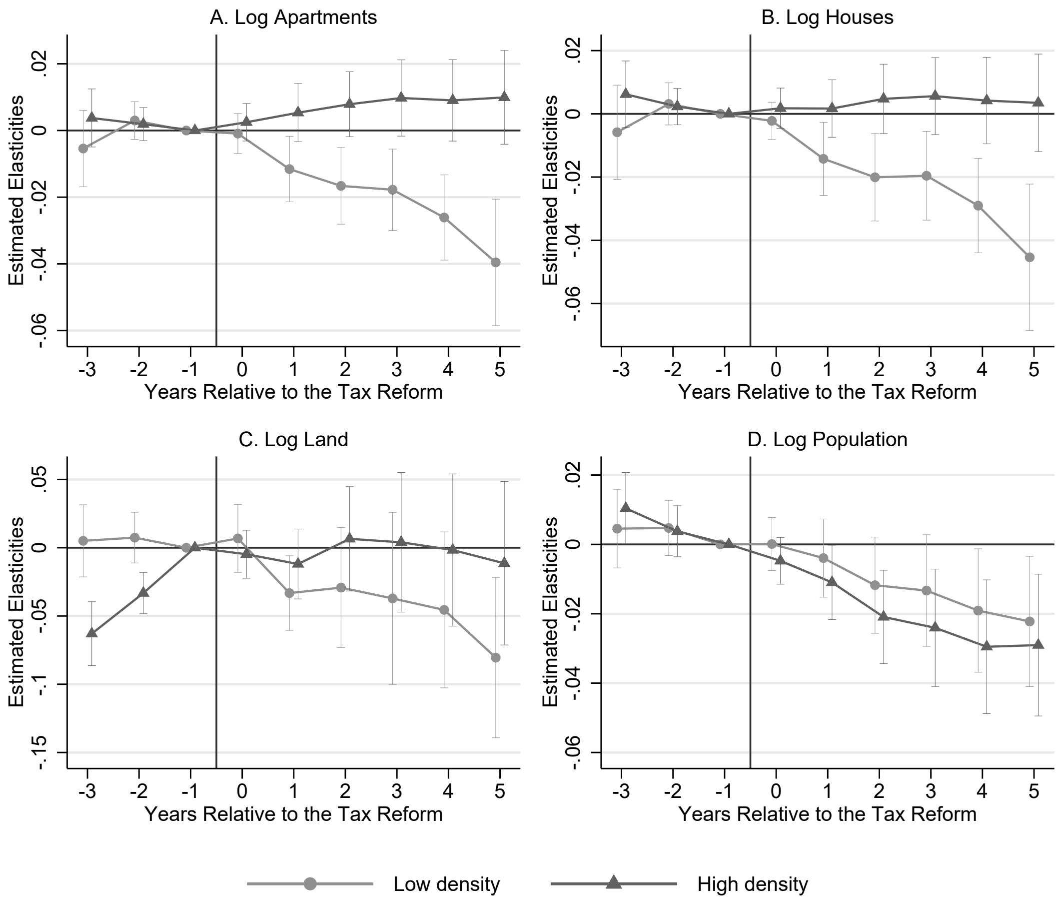
Notes: This figure shows the effects of property taxes on quantities by city size using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.11: The Effects of Property Taxes on Prices by Density



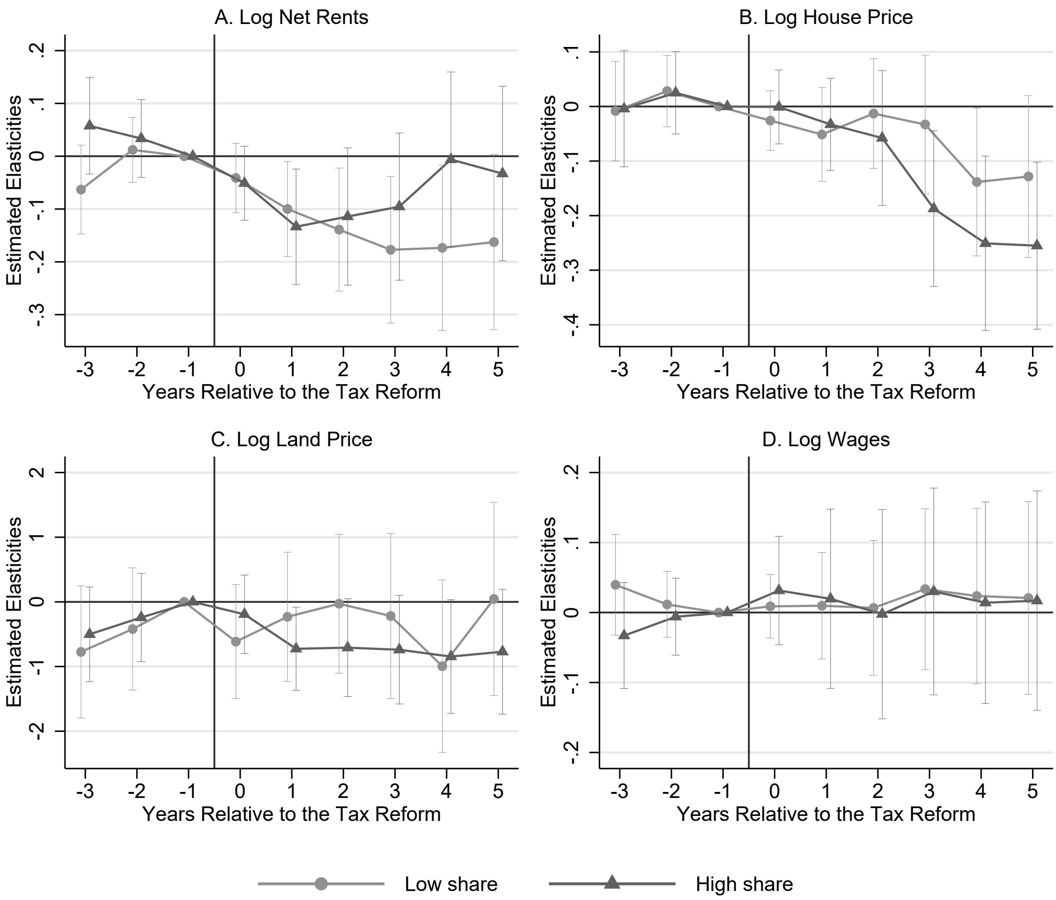
Notes: This figure shows the effects of property taxes on prices by density using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.12: The Effects of Property Taxes on Quantities by Density



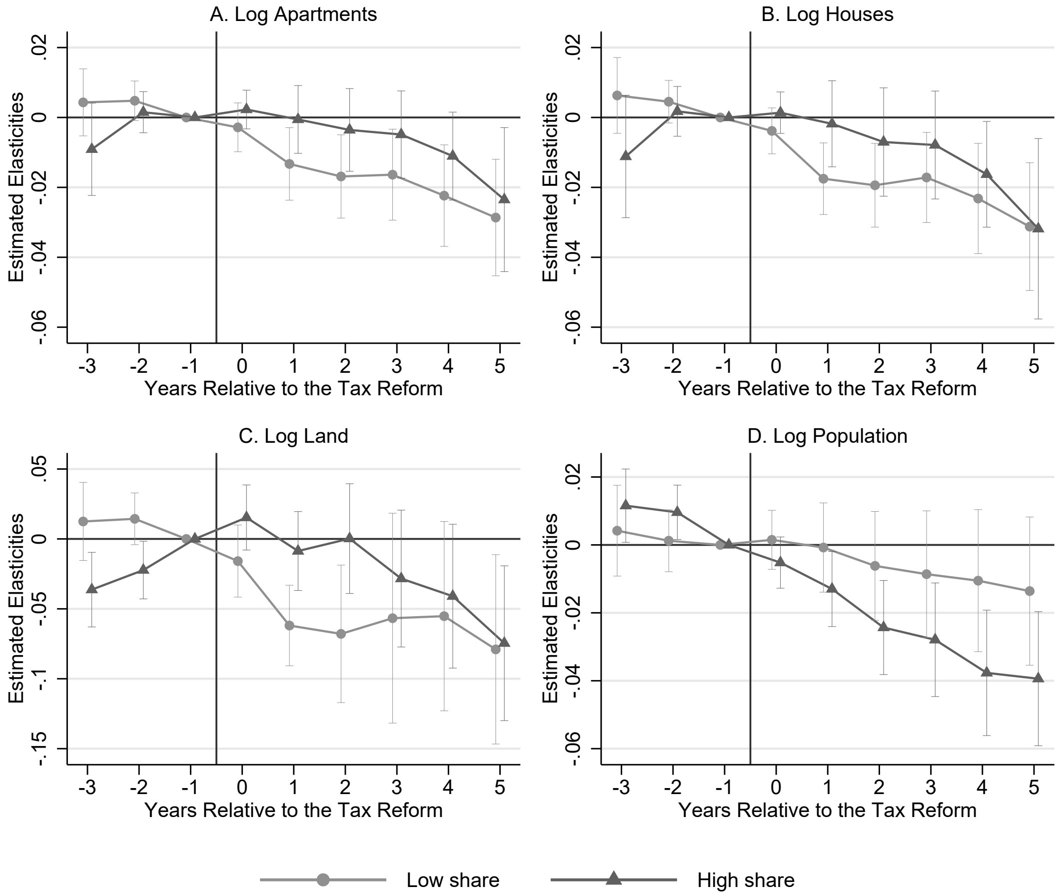
Notes: This figure shows the effects of property taxes on quantities by density using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.13: The Effects of Property Taxes on Prices by Undevelopable Land



Notes: This figure shows the effects of property taxes on prices by undevelopable land using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

Figure 1.C.14: The Effects of Property Taxes on Quantities by Undevelopable Land



Notes: This figure shows the effects of property taxes on quantities by undevelopable land using the event study set-up from equation (1.17). The event is defined as an increase in the local property tax rate (all increases). Standard errors are clustered at the municipality level, vertical bars indicate 90 % confidence intervals. See Table 1.A.1 in the Appendix for details on all variables.

1.C.3 Welfare Effects

Table 1.C.1: Reduced Form Elasticities – Difference-in-Difference Estimates

	Base Sample			Full Sample		
Outcome	Elasticity	S.E.	Obs.	Elasticity	S.E.	Obs.
Log Net Rent	-0.018	(0.032)	37,672	-0.018	(0.032)	37,672
Log House Price	-0.038	(0.031)	33,767	-0.038	(0.031)	33,767
Log Apartments	0.010	(0.004)**	2,780	-0.004	(0.003)	93,192
Log Houses	0.011	(0.005)**	2,780	-0.003	(0.004)	93,192
Log Land Price	-0.070	(0.260)	1,205	0.007	(0.212)	6,987
Log Land	-0.025	(0.014)*	1,712	-0.011	(0.008)	50,825
Log Wage	0.006	(0.027)	973	-0.024	(0.040)	8,067
Log Population	0.000	(0.011)	5,050	-0.001	(0.008)	189,505

Notes: This table summarizes the reduced-form results for the key medium-run elasticities of our model for both the house price sample and the full sample. For detailed information on all variables, see Appendix Table 1.A.1.

Table 1.C.2: Welfare Effects of Property Tax Increases

Housing Share	$\alpha = 0.2$			$\alpha = 0.3$			$\alpha = 0.4$		
Labor Share β	0.45	0.55	0.65	0.45	0.55	0.65	0.45	0.55	0.65
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A – Marginal Welfare Effects									
Worker/Tenant	-0.167	-0.167	-0.167	-0.258	**-0.258**	-0.258	-0.348	-0.348	-0.348
Firm Owner	-0.504	-0.414	-0.325	-0.504	**-0.414**	-0.325	-0.504	-0.414	-0.325
Land Owner	-0.512	-0.512	-0.512	-0.512	**-0.512**	-0.512	-0.512	-0.512	-0.512
Total Welfare Effect	-1.183	-1.094	-1.005	-1.273	**-1.184**	-1.095	-1.364	-1.274	-1.185
Panel B – Welfare Loss Shares									
Worker/Tenant	0.141	0.153	0.166	0.202	**0.218**	0.235	0.255	0.273	0.294
Firm Owner	0.426	0.379	0.324	0.396	**0.350**	0.297	0.369	0.325	0.275
Land Owner	0.433	0.468	0.510	0.402	**0.432**	0.468	0.375	0.402	0.432

Notes: This table presents marginal welfare effects and welfare shares based on the following three reduced form elasticites $d\ln r_c^{H*}/d\ln\tau_c = -.096$, $d\ln l_c^{H*}/d\ln\tau_c = -.512$, $d\ln w_c^{H*}/d\ln\tau_c = .014$ and based on the calibrated housing share in consumption (α) and labor share in production (β) as indicated in the table head. Our preferred specification is given in column (5).

Chapter 2

The Long-Term Costs of Government Surveillance*

2.1 Introduction

More than one third of the world population lives in authoritarian states that attempt to control almost all aspects of public and private life (The Economist Intelligence Unit, 2014). To this end, those regimes create large-scale surveillance systems that infiltrate the population and generate a widespread atmosphere of suspicion reaching deep into private spheres (Arendt, 1951). Such environments of distrust are thought to have adverse economic effects, since they limit cooperation and the open exchange of ideas (Arrow, 1972, Putnam, 1995, La Porta et al., 1997, Algan and Cahuc, 2014). However, systematic empirical evidence on the relationship between government surveillance, trust, and economic performance is missing.

In this paper, we investigate the long-run social and economic consequences of one of the largest and densest surveillance networks of all time. Using administrative data on the ubiquitous network of informers during the times of the socialist German Democratic Republic (GDR), we estimate the long-term, causal effects of government surveillance on measures of trust and economic performance after the fall of the Iron Curtain. The GDR Ministry for State Security, commonly referred to as the *Stasi*, administered a huge body of so-called *Inoffizielle Mitarbeiter*, unofficial informers. These informers accounted for around one percent of the East German population in the 1980s and were regarded as the regime's most important instrument to secure power (Müller-Enbergs, 1996, p. 305). Informers were ordinary citizens who kept their regular jobs but secretly gathered information within their professional and social network, thus betraying the trust of friends, neighbors, and colleagues (Bruce, 2010). A large body of historical research deems the effects of the surveillance apparatus to be devastating and long-lasting: "The oppressive effects of the constant threat of Stasi surveillance [...] can scarcely be overstated.

* This chapter is based on a revised version of A. Lichter, M. Löffler, and S. Siegloch (2016). "The Long-Term Costs of Government Surveillance: Insights from Stasi Spying in East Germany". *CESifo Working Paper* 6042.

It led to perpetual insecurity in personal relationships, and was to leave a difficult legacy for post-reunification Germany" (Fulbrook, 2009, p. 221).[1]

We put these claims to a causal test by exploiting regional variation in the surveillance intensity across East German counties. We explicitly address the concern that recruitment of informers across counties might not have been random by adopting a border discontinuity design that exploits the specific administrative structure of the surveillance state. Stasi district offices bore full responsibility for securing their territory and supervising their respective subordinate county offices, which caused surveillance intensities to differ substantially across GDR districts (Engelmann and Schumann, 1995). Indeed, around 25 percent of the variation in the spying density across counties can be explained by district fixed effects. Important for our identification strategy, this structure was at odds with the fully centralized political system of the GDR, where all political powers were held by the central government (Niemann, 2007). This setting allows us to follow Dube et al. (2010) and use discontinuities in surveillance intensities along district borders as a source of exogenous variation. We corroborate the identifying assumption of no other discontinuities at district borders by testing the smoothness of observable characteristics at borders, finding no systematic differences. In addition, we tighten identification by combining the border design with an instrumental variables approach. Using the leave-out average surveillance intensity at the district level as an instrument, we isolate the part of the variation in the county-level spying density that is explained by differences in surveillance strategies across districts.

Overall, the results of our study offer substantial evidence for negative and long-lasting effects of government surveillance on individuals' trust and economic performance. Using data from the German Socio-Economic Panel (SOEP), we find that a higher spying density leads to lower trust in strangers and stronger negative reciprocity, two standard measures predicting cooperative behavior (Glaeser et al., 2000, Dohmen et al., 2009). A one standard deviation increase in the spying density decreases trust (reciprocal behavior) by 0.06-0.09 (0.15-0.17) of a standard deviation, equivalent to a decrease of about 2 percentage points, respectively. We further observe negative effects of state surveillance on two standard measures of institutional trust (Putnam, 2000, Guiso et al., 2004). Individuals in counties with a higher informer density are less likely to attend elections and more likely to vote for extreme parties (conditional on participating). The effects on trust are accompanied by negative and persistent effects on measures of economic performance. We find that individuals in counties with higher spying densities experience longer and/or more spells of unemployment. A one standard deviation increase in the surveillance intensity raises the overall time spent in unemployment by 0.07-0.11 standard deviations, which is equivalent to an increase of about 2 percentage points. We

[1] See also Gieseke (2014, p. 95) or Childs and Popplewell (1996, p. 111).

further show that more government surveillance significantly reduces individual labor income conditional on working: a one standard deviation increase in the spying density decreases gross labor income by 0.11-0.16 of a standard deviation, or 100-140 EUR per month.

Our empirical results become stronger when tightening identification and moving from cross-sectional OLS estimates to the border design, exploiting within county-pair variation at district borders only. Likewise, we see larger estimates when additionally instrumenting the county-level spying density with the (leave-out) district average, suggesting that endogeneity biases estimates towards zero. In line with this finding, we further show that effects are larger when focusing on those county pairs that were part of the same province during the times of the Weimar Republic but were assigned to different districts when the regime dissolved the provinces after World War II to eliminate the power of all subnational entities. Here, the identifying assumption of no other discontinuities at district borders is even more likely to hold as counties should be more similar in unobservable cultural traits. In addition, we provide a wide range of tests to demonstrate the robustness of our results, such as using different measures of surveillance, an alternative definition of the instrument, or different specifications of the border design. We also rule out that other factors, such as socialist indoctrination, drive our effects. Last, we demonstrate that our inference is robust to clustering standard errors at different levels, including the level of districts while accounting for the small number of clusters using bootstrap procedures as suggested by Cameron et al. (2008).

Our study is closely linked to the steadily growing literature on the relationship between institutions, culture (trust), and economic performance (see Algan and Cahuc, 2014, Alesina and Giuliano, 2015, Fuchs-Schündeln and Hassan, 2016, for recent surveys and Section 2.3 for a more detailed discussion of the literature). We confirm the long-term positive effects of institutional quality on economic performance, highlighting the importance of trust, social capital, and social ties for economic prosperity. Econometrically, we refine current identification strategies to estimate causal effects of formal institutions on culture and economic outcomes by combining within-country variation (Tabellini, 2010, Alesina et al., 2013) with spatial border designs (Becker et al., 2016, Fontana et al., 2017). In contrast to other studies, our identifying variation is not generated by deep, historical differences such as religion, ethnicity, or education, but induced by a rather recent, yet pervasive political experiment.

Moreover, we contribute to the literature investigating the transformation and legacy of countries of the former Eastern Bloc after the fall of the Iron Curtain (see, e.g., Shleifer, 1997). In this regard, two other papers study the effects of Stasi surveillance.[2] An older paper by Jacob and Tyrell (2010) looks at the relationship between surveillance, social capital and economic

[2] In addition, Glitz and Meyersson (2016) exploit information provided by East German foreign intelligence spies in the *West* to investigate the economic returns of industrial espionage.

performance. A second paper by Friehe et al. (2015), which was pursued simultaneously but independently from our project, investigates the effects of Stasi surveillance on personality traits. Both studies document negative effects of government surveillance, which are in line with our findings, given that trust is generally considered a component of social capital, and personality traits are known to shape beliefs such as trust (Dohmen et al., 2009). Our paper explicitly investigates the link between state surveillance and economic performance, highlighting the role of interpersonal trust, which has been identified as an "economic primitive" to explain growth (see Berg et al., 1995, and Section 2.3 for a more detailed discussion of the literature). Econometrically, our study advances the literature by explicitly addressing the non-randomness of the county-level surveillance density. We go beyond cross-sectional correlations by combining a border discontinuity design with an instrumental variables strategy. Our findings show that ignoring the endogeneity of the regional surveillance intensity can lead to a non-negligible bias in the estimates. Last, our paper also complements evidence from Alesina and Fuchs-Schündeln (2007) by showing that the East German regime did not only affect individual preferences for redistribution, but also had long-lasting effects on trust and social behavior.

The remainder of this paper is organized as follows. Section 2.2 presents the historical background, the institutional details of the Stasi surveillance system, and our measure of the regional surveillance intensity. In light of the GDR surveillance state, Section 2.3 lays out our conceptual framework, combining theoretical predictions with empirical insights from the literature on trust and economic performance. Section 2.4 introduces our research design and describes the data used in the empirical analysis. Results are presented in Section 2.5. Section 2.6 concludes.

2.2 The GDR Surveillance State

After Germany's unconditional surrender and the end of World War II in May 1945, the country's territory west of the Oder-Neisse line was divided among the four Allied Forces – the US, the UK, France and the Soviet Union. While the Western forces soon established the principles of democracy and free markets in their respective zones, the Soviet Union implemented a socialist regime in the eastern part of the country. In May 1949, the ideological division of the nation was institutionalized when the Federal Republic of Germany was established on the territory of the three western zones. Five months later, the German Democratic Republic (GDR) was constituted in the Soviet ruled zone, which eventually led to a 40 year long division of the country.

The GDR was an authoritarian regime under the rule of the Socialist Unity Party (SED) and its secretaries general. Its organization closely followed the Soviet example of a highly centralized state, with all political power being held by the Politburo in East Berlin. To this end, the regime abolished existing decentralized political institutions from the times of the Weimar Republic and eliminated the power of subnational entities shortly after the end of World War II. In a first step, the Soviet occupying forces formed the five intermediate jurisdictions Mecklenburg, Anhalt, Brandenburg, Thuringa, and Saxony, which were eventually abolished in 1952 and replaced by 15 administrative districts (*Bezirke*).[3] These districts held no legislative powers. Instead, "the only task [...] was to execute the decisions made by the central committee. This was their raison d'être." (Ulrich Schlaak, Second Secretary of the SED in the district of Potsdam, as cited in Niemann, 2007, p. 198, own translation).

In the early years, the GDR was under constant internal pressure. Dissatisfaction with working conditions and the implementation of socialism culminated in the People's Uprising on and around June 17, 1953, when an unexpected wave of strikes and demonstrations hit the country. Another clear indication of discontent with the regime was the massive and steady out-migration. From 1949 to 1961, roughly 2.7 million citizens (around 20 percent of the population) managed to migrate to the West by authorized migration or illegal border crossing (see Figure 2.A.1). Securing the inner-German border in 1952 was not sufficient to stop this exodus, as people could still escape to the West relatively easily via the divided city of Berlin. Eventually, the regime stopped the substantial population loss by building the Berlin Wall in 1961, and ordering soldiers to shoot at every person trying to illegally cross the inner-German border. Between 1961 and 1988 only around 0.1 percent of the population managed to emigrate (legally or illegally) to the West on an annual basis.

From this point on, East and West Germany increasingly grew apart in their social and cultural patterns. East Germans "felt they had to try to work with socialism, and to confront and make the best of the constraints within which they had to operate" (Fulbrook, 2009, p. 174). This pragmatic attitude led to a situation of relative political stability in the following two decades. On the evening of November 9, 1989, the fall of the Berlin Wall marked the beginning of the dissolution of the GDR, which officially ended with the reunification of West and East Germany in October 1990.

The Ministry for State Security. In February 1950, just a few months after the proclamation of the GDR, the Ministry for State Security was founded. The Stasi served as the internal (and external) intelligence agency of the regime. Its official mission was to "battle against agents, saboteurs, and diversionists [in order] to preserve the full effectiveness of [the] Constitution."[4] Soon after its foundation and the unforeseen national uprising against the regime in June 1953, the Stasi substantially expanded its activities and turned into an ubiquitous institution, spying

[3] Initially, 14 districts were created. In 1961, East Berlin was declared a district of its own.

[4] According to Erich Mielke, subsequent Minister for State Security from 1957 to 1989, on January 28, 1950 in the official SED party newspaper *Neues Deutschland* as quoted in Gieseke (2014, p. 12).

on and suppressing the entire population to ensure and preserve the regime's power (Gieseke, 2014, p. 50ff.).

The key feature of the Stasi's surveillance strategy was the use of "silent" methods of repression rather than legal persecution by the police (Knabe, 1999). To this end, the Stasi administered a dense network of unofficial informers, the regime's "main weapon against the enemy"[5], who secretly gathered detailed inside knowledge about the population. "Informers were seen as an excellent way of preventing trouble before it started [...]" (Childs and Popplewell, 1996, p. 83). In the 1980s, the Stasi listed around 85,000 regular employees and 142,000 unofficial informers, which accounted for around 0.5 and 0.84 percent of the population, respectively.[6]

The organizational structure of the Stasi differed markedly from the otherwise highly centralized political system. While the East German surveillance apparatus was decentralized from the very beginning (Naimark, 1994), the Stasi further increased the responsibilities of its regional offices in the mid-1950s to extract information more efficiently (Engelmann and Schumann, 1995). The Stasi maintained district offices (*Bezirksdienststellen*) in each district capital. These offices bore full responsibility for securing their territory and were independent in how to achieve this goal (Gill and Schröter, 1991, Engelmann and Schumann, 1995, Gieseke, 2014).[7]

As a consequence of this decentralized structure, surveillance strategies differed substantially across GDR districts. Overall, district differences account for a quarter of the variation in informers across counties.[8] The historical literature identifies a small number of hard factors that are believed to have driven differences in surveillance strategies, such as population size, the presence of strategically important firms and industries as well as the strength of the political opposition (Horsch, 1997, Müller-Enbergs, 2008). Besides these systematic drivers, other, non-systematic factors are thought to have played a role as well. These soft and arguably more random determinants include effort, zeal or the loyalty of the district leadership to the regime (Gill and Schröter, 1991, Childs and Popplewell, 1996, Müller-Enbergs, 2008). We exploit these insights to develop our identification strategy in Section 2.4.1.

[5] Directive 1/79 of the Ministry for State Security for the work with unofficial collaborators (Müller-Enbergs, 1996, p. 305).

[6] The number of regular Stasi employees was notably high compared to the size of other secret services in the Eastern Bloc (see, e.g., Albats, 1995, Gieseke, 2014, p. 72, and Harrison and Zaksauskiene, 2015).

[7] The Minister for State Security hardly influenced the activities and directives governed by the heads of the district offices (Gill and Schröter, 1991). Moreover, according to various accounts, the Politburo did not exert any control over the Ministry for State Security from the mid-1950s onwards (see Childs and Popplewell, 1996, p. 67).

[8] Similarly, there were sharp differences in other domains of the surveillance system. For instance, around one third of the constantly-monitored citizens (*Personen in ständiger Überwachung*) were living in the district of Karl-Marx-Stadt (Horsch, 1997), which accounted for only eleven percent of the total population. Likewise, 17 percent of the two million bugged telephone conversations were tapped in the district of Magdeburg, which only made up eight percent of the population.

Unofficial Informers. Each district office had authority over the county offices (*Kreisdienststellen*) and on-site offices (*Objektdienststellen*) within their territory.[9] In total, there were 209 county offices, which executed the commands and orders from their respective district office and recruited and administered the Stasi informers. These informers were instructed to secretly collect information about individuals in their own network. After recruitment it was thus necessary for informers to pursue their normal lives as friends, colleagues, and neighbors. The Stasi administered the body of informers in a formalized way, with cooperation being sealed in written agreements and informers being closely led by a responsible Stasi officer (Gieseke, 2014, p. 114ff.). Informers would secretly meet with their officer on a regular basis to report suspicious behavior and provide personal information about individuals in their social networks. Reasons for serving as a collaborator were diverse. Some citizens agreed to cooperate for ideological reasons, others were attracted by the personal and material benefits that accompanied their cooperation. However, in rare cases the regime also compelled citizens to act as unofficial informers by creating fear and pressure (Fulbrook, 2005, p. 242f.). With the collected intelligence at hand, the Stasi was able to draw a detailed picture of anti-socialist and dissident movements within the society and to exert an overall "disciplinary and intimidating effect" on the population (Gieseke, 2014, p. 84f.).

Numerous historical accounts suggest that the population was aware of the large network of informers. "The very knowledge that the Stasi were there and watching served to atomize society, preventing independent discussion in all but the smallest groups" (Popplewell, 1992). The consequence was "the breakdown of the bonds of trust between officers and men, lawyers and clients, doctors and patients, teachers and students, pastors and their communities, friends and neighbors, family members and even lovers" (Childs and Popplewell, 1996, p. 111). The preferred method of the Stasi was "to build up and propagate distorted stories with enough kernel of truth to sow suspicion and discredit the individual" (Fulbrook, 1995, p. 54), eventually destroying relationships, reputations, and careers.[10]

Measuring Surveillance Intensity. Given that the Stasi saw unofficial collaborators as their main instrument of surveillance, we choose the county-level share of informers in the population as our preferred measure of surveillance intensity. Although the Stasi was able to destroy parts

[9] On-site offices were separate entities in seven strategically important public companies or universities. The Stasi only monitored economic activity but was not actively involved in economic production (Gieseke, 2014).

[10] Bruce, 2010 reports that the vast majority of people had direct contact with the Stasi multiple times throughout their lives; Fulbrook, 1995 states that friendships inevitably had a shadow of distance and doubt; Wolle, 2009 writes that the threat of being denounced caused an atmosphere of mistrust and suspicion within a deeply torn society; and Reich, 1997 more figuratively describes that citizens felt the Stasi's presence like a "scratching T-shirt". For less scientific, more popular representations of the impact of the Stasi, see the Academy Award-winning movie "The Lives of Others" and the TED talk "The Dark Secrets of a Surveillance State" given by the director of the Berlin-Hohenschönhausen Stasi prison memorial, Hubertus Knabe.

of its files in late 1989, much of the information was preserved when protesters started to occupy Stasi offices across the country. In addition, numerous shredded files have been restored since reunification by the Stasi Records Agency (BStU) – a government agency established in 1990/1991 to safe-keep and secure the Stasi Records and to provide people, researchers, and media access to these files. Our data on the number of unofficial informers in each county are based on these official records. Most of the data have been compiled in Müller-Enbergs (2008). Until today, the Stasi Records Agency keeps restoring old files and releasing new data and information. Hence, we extended the data in Müller-Enbergs (2008) with newer data which we collected from the archives of the Agency. Overall, this allows us to observe the spying density for around 92 percent of the counties at least once in the 1980s.[11]

The Stasi officially differentiated operative collaborators (*IM1*) from collaborators providing logistics (*IM2*).[12] Our baseline measure of the county-level surveillance density is based on the operative collaborators as these informers (i) were actively involved in spying, (ii) constituted the largest and most relevant group of collaborators, and (iii) exhibit the best data coverage across counties.[13] If an on-site office was located in a county, we add the respective number of unofficial informers to the county total and explicitly control for the presence of these on-site offices in the econometric analysis. As information on the total number of collaborators is not given for each year in every county, we use the average share of informers from 1980 to 1988 as our measure of surveillance. The spying density was stable across the 1980s, the within-county correlation being 0.91. For further details on our main explanatory variable, see Data Appendix 2.B.1.

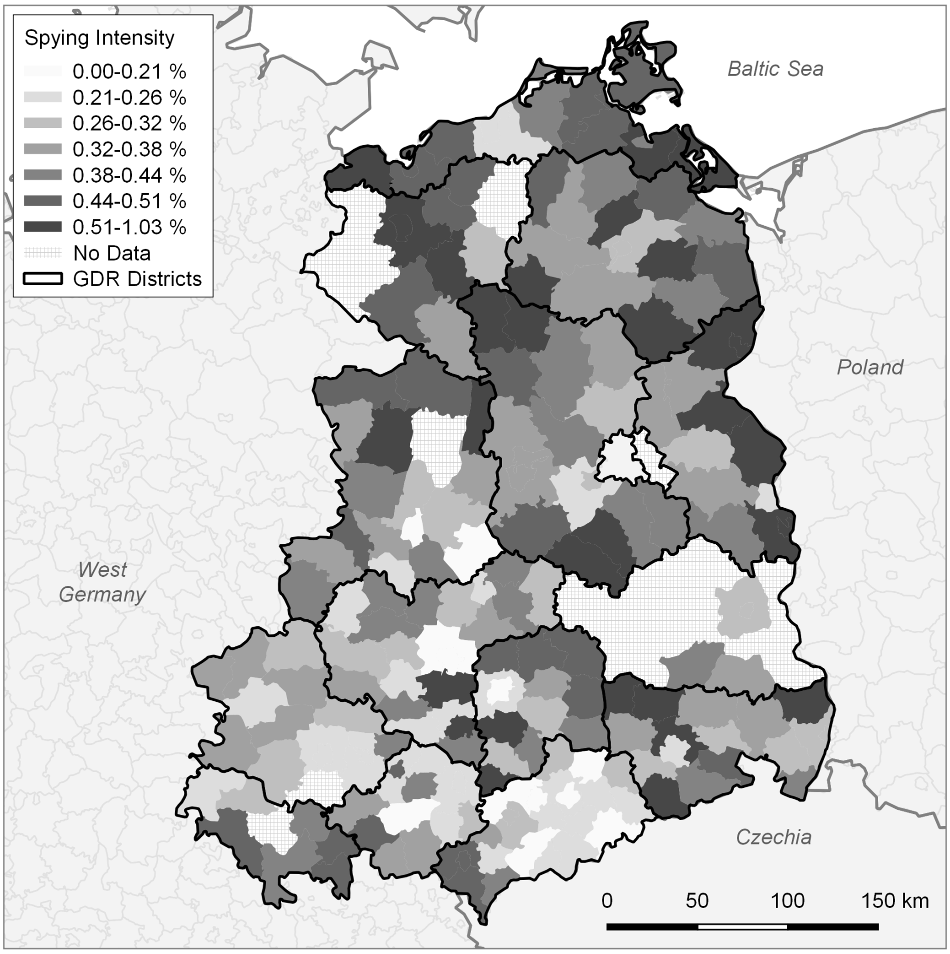
Hence, our baseline measure of county-level surveillance is the spying density defined as the number of operative informers divided by the county population in the respective year. As informers were the central weapon of the system, this measure is arguably the best proxy to pick up the effect of the Stasi surveillance apparatus as a whole. By definition, this overall effect also comprises the specific *modus operandi* of the Stasi, i.e., using informers within social networks. Likewise, our measure should also account for more persecutive activities such as arrests. While such measures only played a minor role in the strategy of the Stasi, the regime occasionally had to make use of physical violence and imprisonment to lend credibility to the overall surveillance system.

[11] Pre-1980 data are only available for a limited number of counties.

[12] In some of the Stasi's informer accounts, there is a third category called "societal collaborators". These individuals were publicly known to be loyal to the regime and in the vast majority of cases not involved in spying. Rather, these collaborators were asked to actively and openly support the Stasi and the regime (Kowalczuk, 2013). In this sense, they were less secret than official Stasi employees who oftentimes disguised their connections to the regime.

[13] Nonetheless, we show that results are robust when combining IM categories 1 and 2, hence using the total number of spies as the main regressor.

Figure 2.1: Spying Intensity across Counties



Notes: The figure shows the county-level surveillance density measured by the average yearly share of operative unofficial informers relative to the population between 1980 and 1988. *Source:* Data on the number of unofficial informers is taken from Müller-Enbergs (2008) and additionally collected from the Stasi Records at the BStU, see Data Appendix 2.B.1 for details. *Maps:* MPIDR and CGG (2011) and EuroGeographics.

Figure 2.1 plots the regional variation of surveillance intensity, darker colors indicating higher spying densities. The surveillance intensity differs considerably both across and within GDR districts. The share of operative informers in a county ranges from 0.12 to 1.03 percent, the mean density being 0.38 percent (see Table 2.B.2 for more detailed distributional information). The median is similar to the mean (0.36 percent), and one standard deviation is equal to 0.14 informers per capita. In our regressions, we standardize the share of informers by dividing it by the respective sample standard deviation.

2.3 Conceptual Framework and Related Literature

Theoretically, we expect Stasi surveillance to have long-term negative effects on economic performance through a persistent deterioration of interpersonal and institutional trust. Authoritarian regimes establish a system of obedience by instilling fear (Arendt, 1951) and generate uncertainty in the rewards of individual productive investments, eventually leading to lower economic activity and a negative effect on growth (Smith, 1776). In line with this rationale, a large literature has shown the positive association between the quality of political institutions and economic prosperity (see, e.g., Mauro, 1995, Hall and Jones, 1999, Rodrik et al., 2004).

In this context, lacking trust has been identified as one key factor for linking poor political institutions and low economic performance. From a theoretical perspective, trust is assumed to be essential for growth as every economic transaction involves an element of mutual confidence (Arrow, 1972). In this spirit, trust works as social collateral. It reduces transaction costs by limiting the need for formal institutions to enforce agreements, and triggers investment and other economic activity as policy pronouncements appear more credible (Knack and Keefer, 1997). This role of trust as an "economic primitive" has been documented in various studies in experimental economics, which demonstrate that it fosters reciprocal behavior and cooperation, and thereby leads to more efficient economic outcomes (see, e.g., Berg et al., 1995, Dohmen et al., 2008, 2009).

Various non-experimental studies have investigated and confirmed this link between trust and economic performance as well.[14] Knack and Keefer (1997) and Zak and Knack (2001) document a positive correlation between trust and economic indicators across countries. In two related papers, Nunn (2008) and Nunn and Wantchekon (2011) show that transatlantic and Indian Ocean slave trade still has a causal and persistently negative effect on current trust levels and economic performance in Africa. Algan and Cahuc (2010) isolate the trust that US descendants have inherited from their forebears who immigrated from different countries at different dates and show that time variation in inherited trust impacts economic growth in the respective countries of origin. In a series of papers, Guiso et al. (2006, 2009) exploit variation in deep cultural aspects, such as religious affiliation, to explain trust levels, which in turn have real economic effects. Similarly, Tabellini (2010) exploits variation in literacy and political institutions at the end of the 19th century to explain trust levels and regional economic development across European countries in the 1990s.

Last, there is evidence that even temporary changes in political institutions can have long-

[14] In lieu of trust, some of the these papers use the term social capital, initially coined by political scientists (Putnam, 1993, Fukuyama, 1995). While exact definitions of social capital vary, most of them include trust, cooperative norms, and association within groups as key components (Knack and Keefer, 1997). As a result, we use the terms trust and social capital interchangeably.

term effects. A crucial element for the legacy of short-lived changes in political institutions is the intergenerational transmission of attitudes. An important literature, drawing on the work of Bisin and Verdier (2000) and Bisin et al. (2004), has shown that cooperative values (Tabellini, 2008) and cultural beliefs (Guiso et al., 2008) are transmitted from one generation to the next. Dohmen et al. (2012) confirm this hypothesis empirically using direct survey measures. Specifically, they show that trust and reciprocity are transmitted from parents to children and confirm the implicit assumption of Algan and Cahuc (2010). In light of this intergenerational transmission of attitudes and beliefs, several studies provide further evidence for persistent economic effects of temporary shocks to political institutions. Acemoglu et al. (2001) study the long-term effects of differences in colonial institutions; La Porta et al. (1998) assess the long-run effects of financial legislation; Nunn (2008) and Nunn and Wantchekon (2011) demonstrate the legacy of slave trade in Africa. Based on these insights, we expect the Stasi to have a persistent negative effect on trust and social capital, even after the end of the GDR regime and Germany's reunification.

2.4 Research Design and Data

In the following, we first set up the empirical model and lay out the research design (Section 2.4.1). In Section 2.4.2, we describe the data used for the empirical analysis. Last, we provide evidence that covariates are smooth at district borders, corroborating the main identifying assumption of our empirical design (Section 2.4.3).

2.4.1 Empirical Model and Identification

Our identification strategy exploits the administrative structure of the Stasi, where district offices bore the full responsibility for securing their territory and administered different average levels of the informer density at the county level. As a result, district fixed effects explain more than a quarter of the county-level variation in the spying density. We harness the resulting discontinuities along district borders as a source of exogenous variation and set up a standard border design (see, e.g., Holmes, 1998, Dube et al., 2010, Magruder, 2012, for studies applying similar research designs).

Limiting the analysis to contiguous counties that straddle a GDR district border, we formally regress outcome Y (see Section 2.4.2 for a detailed list of outcomes) of individual i in county c, which is part of a border county pair b and situated in a former Weimar province p, on the spying density in county c and county-pair dummies ν_b:

$$Y_{icbp} = \alpha + \beta \times SPYDENS_c + \boldsymbol{X_i}\boldsymbol{\delta} + \boldsymbol{K_c}\boldsymbol{\phi} + \nu_b + \mu_p + \varepsilon_{icbp}. \tag{2.1}$$

In addition, our baseline model includes sets of covariates at the individual and county level, denoted $\boldsymbol{X_i}$ and $\boldsymbol{K_c}$, respectively. At the individual level, we control for the age and gender of the respondents. County-level covariates account for the above mentioned systematic factors determining the surveillance strategy and include controls for differences in size, industrial composition, oppositional forces, and long-run political differences measured pre World War II (see Section 2.4.2 for a detailed description of control variables). We also include a set of dummy variables indicating pre-World War II provinces from the Weimar Republic, denoted μ_p, which captures long-term cultural differences, e.g., between Prussia and Saxony (see Figure 2.A.4 for a mapping of GDR districts into provinces from the times of the Weimar Republic). In addition, all regressions include a dummy for the presence of an on-site office (cf. Section 2.2). The error term is denoted by ε_{icbp}.

Identification. Equation (2.1) exploits variation within county pairs that is due to county characteristics and district-level surveillance strategies. The identifying assumption is that a given county on the lower-spying side of a district border is similar to its neighboring county on the higher-spying side in all relevant characteristics except for the spying density. If this is fulfilled, the remaining source of variation is district-level variation in surveillance and our estimates produce the causal effect of the spying density. We test the smoothness of observable county characteristics at district borders in Section 2.4.3 using a standard covariate-smoothness test. We find no significant differences in covariates, corroborating the identifying assumption. However, other potential threats to identification still exist.

First, unobservable confounders within county pairs might drive parts of the county-level spying density. To account for this potential source of endogeneity, we estimate another specification in which we isolate the district component of surveillance within county pairs by replacing the county-level spying density with the leave-out surveillance intensity at the district level.[15]

Second, selection effects pre and post-reunification could invalidate our research design. While out-migration was very limited after the construction of the Berlin Wall (cf. Section 2.2), residential mobility *within* the GDR was also highly restricted as all living space was tightly administered and allocated by municipal housing agencies (Grashoff, 2011, p. 13f.). Post-reunification, we assign treatment based on the county of residence in 1989, and follow individuals over time, also when they change residency.[16]

[15] The leave-out district average for county *A* is the average of all counties in the district *excluding* the spying density of county *A*. Results are similar when using the simple mean district-level spying density.

[16] We additionally check for differences in treatment effects for individuals that stayed or moved after the fall of the Berlin Wall. Estimates are very similar for both groups, which suggests that effects of surveillance persist even if individuals change their residence and move to other counties. Given that any mobility response post-

Third, our proxy of surveillance intensity may not translate into differences in awareness of the Stasi. While various historical accounts suggest that the population was well aware about state surveillance (cf. Section 2.2), we can also test this assumption. Since 1992, any citizen can file a request to view his or her personal Stasi file. Based on official data on the total number of individual requests for disclosure by district and year obtained from the Stasi Records Agency, we find a strong positive correlation between the per-capita number of individual requests in a district and the corresponding district-level spying density (see Panel A of Appendix Figure 2.A.2) as well as a rather stable pattern over time (see Panel B of Appendix Figure 2.A.2).[17] Moreover, we could face measurement error in the main regressor if (i) informers recruited by one county collected information on individuals located in the neighboring county within the same county pair, or (ii) there was a quantity-quality trade-off in terms of unofficial collaborators. While we cannot rule out these mechanisms, both would work against finding large effects and bias our estimates towards zero.

Last, there may be alternative channels through which the regime could have affected its citizens. If these channels showed discontinuities at district borders and were highly correlated with the spying density at the county level, we might wrongly attribute the negative effects on trust and economic performance to the surveillance system. Given that districts themselves had no legislative power, the most likely candidate in this respect might be socialist indoctrination, which has been shown to affect individual preferences and economic outcomes (see, e.g., Alesina and Fuchs-Schündeln, 2007, Fuchs-Schündeln and Masella, 2016). We proxy socialist indoctrination by the share of SED party members among the political and economic elites in 1988 (see Appendix Table 2.B.1 for details on this variable). We find that this measure is only weakly correlated with the intensity of surveillance (see Figure 2.A.5 in the Appendix) and that estimates remain essentially unchanged when including this variable as a control (see Appendix Table 2.A.5).

Defining Border County Pairs. As shown in Figure 2.1, several counties face more than one direct neighbor on the other side of a district border, a feature that is common in border discontinuity designs.[18] We follow the influential paper by Dube et al. (2010) and duplicate a county subject to the number of bordering counties. However, this approach might overstate

reunification may in itself be driven by the spying density, these findings should, however, not be interpreted causally.

[17] Information on the number of requests is not available at the county level. Results shown in Figure 2.A.2 are robust to the inclusion of controls. One reason for the observed drop in 2003 might be the fact that the Stasi Records Agency started to reconstruct shredded files and made them available to the public from 2004 onwards. Figure 2.A.3 plots the annual number of requests filed between 1992–2012.

[18] We discard East Berlin from our analysis as it was both a county and district on its own. Hence, we cannot construct the leave-out district-level average spying intensity as an instrument.

the importance of pairs that only share a very small borderline but are different in size or other relevant characteristics. For this reason, we also propose an alternative specification, in which we only consider the neighbor that shares the longest border with the respective county. This will be our preferred specification throughout the paper. As this treatment of border counties still leads to the duplication of some counties[19], we account for this by adjusting standard errors for clustering (see next paragraph) and dividing individual weights by the number of duplications. We show below that results are not driven by using different definitions of county pairs (see Appendix Table 2.A.4).

Inference. In our baseline specification, standard errors are two-way clustered at the county-pair and county level to (i) allow for shocks affecting county pairs, and (ii) account for the duplication of some counties in our preferred specification. We provide a more detailed discussion of alternative ways to calculate standard errors, such as clustering at the district level, and demonstrate the robustness of our inference in Section 2.5.2.

2.4.2 Data

To estimate the effect of surveillance on trust and economic performance, we use the German Socio-Economic Panel Study, a longitudinal survey of German households (Wagner et al., 2007, SOEP, 2015). Established for West Germany in 1984, the survey covers respondents from the former GDR since June 1990. This allows us to assign treatment (i.e., the spying density) based on the respondents' county of residence in the year before the fall of the Berlin Wall and follow these respondents over time. Hence, we are able to track individuals even when they changed residence post reunification.

Outcomes. We proxy individual trust and cooperative behavior by two standard variables provided in the SOEP: trust in strangers as specified in Glaeser et al. (2000), and reciprocal behavior (Dohmen et al., 2009). We further operationalize institutional trust by two widely used measures: participation in federal elections and extreme political orientation (Putnam, 2000, Guiso et al., 2004). All trust outcomes are observed in two waves (see Tables 2.B.1 and 2.B.2 for more information); we pool observations from both waves and add year fixed effects.

We further focus on three measures of economic performance. First, we use log mean labor income between 1992–2010. Second, we calculate the total unemployment duration of

[19] To give an example, county *A* in district 1 has two neighbors: county *B* and county *C*, both part of district 2. *A* shares the longest border with *B*, but *C* has only one neighbor, namely *A*. Hence, there are two county pairs *A*-*B* and *A*-*C*, and *A* is duplicated once.

individuals over this period. Third, we calculate the individual probability to become self-employment within the survey period.

Controls. Vectors $\boldsymbol{X_i}$ and $\boldsymbol{K_c}$ in equation (2.1) denote the sets of control variables at the individual and county level, respectively. At the individual level, we control for the respondents' age and gender.[20] At the county-level, we construct four different sets of control variables. First, we account for the *size and demographic composition* of the counties in the 1980s. The corresponding set of controls comprises (i) a county's surface area (in logs), (ii) the log mean county population 1980–1988, (iii) the share of children as of September 30, 1989, and (iv) whether the county is rural or urban (*Stadt-/Landkreis*). Second, we account for differences in the *industrial composition* of counties. The set of industry controls comprises (i) the 1989 share of employees in the industrial sector and the share in the cooperative sector, and (ii) the goods value of industrial production in 1989 (in logs). Third, we control for the strength of *opposition to the regime*. As mentioned in Section 2.2, the uprising in June 1953 constituted the most prominent rebellion against the regime before the large demonstrations in late 1989. The riot markedly changed the regime's awareness of internal conflict and triggered the expansion of the Stasi's informer network. We use differences in the regional intensity of the uprising to proxy the strength of opposition. Specifically, we construct six indicators measuring the strike intensity with values "none", "strike", "demonstration", "riot", "liberation of prisoners", and "military intervention" (no strike serves as the baseline). Fourth, we account for historical and potentially persistent county differences in terms of economic performance and political ideology. The set of *pre-World War II* controls comprises (i) the mean Nazi and Communist vote shares in the federal elections of 1928, 1930, and the two 1932 elections to measure the level of political extremism (Voigtländer and Voth, 2012), (ii) the regional share of Jews and Protestants in 1925 in order to control for religious differences (Becker and Wößmann, 2009), and (iii) the share of white-collar and self-employed workers in 1933 as a proxy for persistent productivity differences across local labor markets.

Summary Statistics. Summary statistics for all outcomes and controls on the individual and county level are presented in Table 2.B.2 in the Appendix.

2.4.3 Covariate Smoothness

The identifying assumption of the border design requires that covariates do not vary systematically at district borders. To test this assumption, we apply a standard smoothness test following

[20] We abstain from controlling for additional covariates at the individual level such as marital status, household size or education as these variables might have been shaped by state surveillance.

Lee and Lemieux (2010), separately regressing each pre-determined control variable on the county-level spying density. This provides a direct test of whether each covariate is unrelated to the informer density.

Table 2.1 reports the results of this test. In column (1), we investigate the smoothness of covariates in the full sample of all East German counties covered by the SOEP. We see that the spying density is significantly correlated with most of our covariates. In column (2), we restrict the sample to counties at district borders but do not include county-pair fixed effects. Hence, we still compare counties which are far away from each other and might differ in many other dimensions than the spying intensity. Again, we detect systematic differences in observables which are correlated with the spying density. In columns (3)–(6), we implement variants of our border design by introducing county-pair fixed effects, hence explicitly testing the smoothness of covariates within county pairs at district borders. In column (3), we consider all possible county pairs, taking into account that counties can have multiple neighbors by duplicating observations accordingly. We do not find significant differences in covariates anymore. The same is true for our preferred specification reported in column (4), in which each county is assigned to the neighboring county with the longest border. In this specification, county pairs that share only a small segment of a district border, and thus might not be ideal matches, are excluded. While individual covariates are balanced in both models, the F-test of joint significance of all covariates, reported at the bottom of Table 2.1, indeed suggests that our baseline model performs better in jointly balancing covariates.[21] Last, we re-estimate models (3) and (4) while additionally accounting for potentially persistent historical differences between the provinces of the Weimar Republic (as we do in the estimations, see equation (2.1)). As discussed before, these provinces held large political power prior to World War II and might have persistently shaped the different regions of the country. Covariates remain balanced, both in the full sample of county pairs (column (5)) and in our preferred sample using those pairs with the longest border only (column (6)).

2.5 Empirical Results

In the following section, we present our empirical findings. Section 2.5.1 presents the main results. In Section 2.5.2, we provide a range of tests to demonstrate the robustness of our results.

[21] The test, suggested by Lee and Lemieux (2010), assesses if some of the discontinuities are statistically significant by random chance, which may happen if the set of covariates is large. While we do not find any significant discontinuity, we nevertheless report the test to assess the overall balancing properties of our estimation samples. Precisely, we test the null hypothesis that all coefficients are jointly equal to zero in a stacked regression.

Table 2.1: Covariate Smoothness at GDR District Borders

	Full Sample (1)	County Pairs (2)	County Pairs (3)	Longest Border (4)	County Pairs (5)	Longest Border (6)
Log Mean Population 1980–1988	-0.809***	-0.430***	-0.282	-0.181	-0.232	-0.107
	(0.166)	(0.122)	(0.175)	(0.192)	(0.185)	(0.222)
Log County Size	0.352***	0.253**	0.007	0.013	-0.000	0.064
	(0.088)	(0.111)	(0.137)	(0.132)	(0.148)	(0.169)
City County	-0.462***	-0.084	0.006	0.007	0.006	0.008
	(0.148)	(0.085)	(0.008)	(0.010)	(0.009)	(0.013)
Share of Population Aged under 15, 1989	0.413***	0.423***	0.092	0.014	-0.004	-0.072
	(0.088)	(0.133)	(0.191)	(0.209)	(0.190)	(0.205)
Log Industrial Output 1989	-0.519***	-0.426***	-0.189	-0.100	-0.166	-0.094
	(0.111)	(0.156)	(0.191)	(0.208)	(0.201)	(0.223)
Share Industrial Employment 09/1989	-0.213**	-0.343**	0.003	0.041	0.048	0.068
	(0.097)	(0.148)	(0.162)	(0.187)	(0.185)	(0.209)
Share of Cooperative Workers 09/1989	0.462***	0.426***	0.036	-0.024	-0.032	-0.075
	(0.092)	(0.136)	(0.156)	(0.155)	(0.170)	(0.176)
Uprising Intensity 1953: None	0.127	0.049	-0.121	-0.147	-0.137	-0.212
	(0.082)	(0.141)	(0.200)	(0.203)	(0.222)	(0.215)
Uprising Intensity 1953: Strike	-0.015	-0.156	-0.166	-0.119	-0.132	-0.071
	(0.104)	(0.103)	(0.123)	(0.117)	(0.149)	(0.147)
Uprising Intensity 1953: Demonstration	-0.017	-0.153	-0.095	-0.129	-0.068	-0.058
	(0.083)	(0.140)	(0.184)	(0.224)	(0.141)	(0.153)
Uprising Intensity 1953: Riot	0.047	0.149	0.190	0.128	0.229	0.207
	(0.113)	(0.127)	(0.196)	(0.218)	(0.206)	(0.230)
Uprising Intensity 1953: Prisoner Liberation	-0.196	0.144	0.282	0.341	0.177	0.201
	(0.172)	(0.129)	(0.220)	(0.243)	(0.181)	(0.203)
Uprising Intensity 1953: Military Intervention	0.027	0.199	0.222	0.269	0.235	0.363
	(0.078)	(0.121)	(0.226)	(0.230)	(0.244)	(0.223)
Mean Vote Share Communist Party 1928–1932	-0.423***	-0.328***	-0.014	0.005	0.007	-0.009
	(0.112)	(0.122)	(0.134)	(0.169)	(0.134)	(0.150)
Mean Vote Share Nazi Party 1928–1932	0.286**	0.215*	-0.040	-0.041	-0.036	-0.091
	(0.110)	(0.109)	(0.200)	(0.226)	(0.171)	(0.183)
Share Protestants 1925	0.158***	0.270**	0.104	0.032	0.100	-0.038
	(0.044)	(0.118)	(0.147)	(0.181)	(0.148)	(0.193)
Share Jews 1925	-0.617*	-0.049	0.105	0.128	0.170	0.218
	(0.333)	(0.168)	(0.200)	(0.239)	(0.192)	(0.228)
Mean Share of Self-Employed 1925 and 1933	0.376***	0.216	0.174	0.106	0.203	0.187
	(0.094)	(0.139)	(0.151)	(0.166)	(0.132)	(0.156)
Share of White Collar Workers 1933	-0.563***	-0.041	0.073	0.057	0.170	0.199
	(0.184)	(0.169)	(0.219)	(0.256)	(0.188)	(0.202)
County-Pair Fixed Effects			Yes	Yes	Yes	Yes
Weimar Province Fixed Effects					Yes	Yes
Counties	148	80	80	80	80	80
County Pairs		72	72	53	72	53
Joint F-Test	8.635	3.887	1.758	1.537	1.228	0.936
p-value	0.000	0.000	0.049	0.115	0.264	0.543

Notes: This table presents the results of our covariate smoothness test. In column (1), we separately regress each covariate on the spying density using the full set of counties. Specification (2) is based on the border county-pair sample. Column (3) adds the set of county-pair fixed effects. In column (4), we further restrict the sample to those county pairs with the longest border in case a county is part of multiple pairs (our baseline). In columns (5) and (6), border definitions are identical to specifications (3) and (4), respectively, but we additionally control for persistent differences across Weimar Provinces by including the corresponding set of fixed effects. All variables have been standardized. Population weights are adjusted for duplications of counties that are part of multiple county pairs. Standard errors are two-way clustered at the county and county-pair level. Significance levels are * $p < .1$, ** $p < .05$, *** $p < .01$. The reported F-test statistics and the corresponding p-values test the null hypothesis of all coefficients being jointly equal to zero in a stacked regression (Lee and Lemieux, 2010). For information on all variables, see Appendix Table 2.B.1.

2.5.1 Main Results

In this section, we analyze the effect of spying on various measures of trust and economic performance, applying the border design as set up in equation (2.1). Tables 2.2 and 2.3 summarize our main findings. In order to demonstrate the relevance of our identification strategy, we implement our research design step-by-step and start with a naive OLS estimate, capturing the correlation between the spying density and our outcomes in the full sample using all counties. In a second step, we restrict the sample to counties located at district borders. While this subsample only contains counties straddling a district border, we do not exclusively compare direct neighbors. In the next step, we implement the border design by including county-pair fixed effects (column (3)). Here, identification is based on differences *within* county pairs at district borders. Specification (4) is our baseline model as described in equation (2.1), including individual and county-level controls to account for potential differences in observable characteristics between counties within county pairs. In a last step, we combine the border design with the specific institutional feature that Stasi district offices administered different average levels of surveillance intensities (see our discussion in Section 2.2). Econometrically, we use the leave-out average district-level surveillance intensity as an instrument for the county-level spying density and estimate the reduced form relationship, reported in column (5). In Appendix Table 2.A.6, we additionally document the corresponding first and second-stage relationships and provide IV-diagnostics. Overall, the instrument proves to be reasonably strong with first-stage F-statistics exceeding 10 for all outcomes and second-stage estimates being statistically significant.[22]

Overall, Table 2.2 shows significantly negative effects of the spying density on our measures of trust in the border design.[23] Importantly, we find that our results become stronger when implementing the identification strategy step-by-step, i.e., when moving from column (2) to column (5). This empirical pattern suggests that endogeneity induces an upward bias in the estimates. We consider this an important insight in light of the current literature using the same variation in government surveillance (Jacob and Tyrell, 2010, Friehe et al., 2015). In terms of magnitudes, we find that a one standard deviation increase in the spying density reduces trust in strangers by 0.07-0.09 of a standard deviation (Panel A, columns (4) and (5)). Reciprocal behavior decreases by 0.15-0.17 of a standard deviation (Panel B). Turning to our two measures of institutional trust, we observe a very similar pattern (cf. Table 2.2, Panels C and D). We find a significant negative effect on the intention to vote in federal elections. A one standard deviation increase in the spying density decreases the likelihood of attending elections by 0.06-0.12 of a

[22] We find similar effects when including each county's own contribution to calculate the district-level spying density, that is, using the overall district average as an instrument (cf. Appendix Table 2.A.6).

[23] We obtain similar results for two alternative measures of individual and institutional trust, the number of close friends and engagement in local politics, see Appendix Table 2.A.1.

Table 2.2: The Effect of Spying on Interpersonal and Institutional Trust

	All Counties	Border County Sample			
	(1)	(2)	(3)	(4)	(5)
Panel A – Trust in Strangers					
County-Level Spying Density	0.051*	0.069*	-0.030	-0.065**	
	(0.029)	(0.039)	(0.041)	(0.027)	
District-Level Spying Density					-0.092**
					(0.044)
Number of Observations	3,307	2,007	2,007	2,007	2,007
Adjusted R-Squared	0.006	0.022	0.109	0.143	0.143
Panel B – Reciprocal Behavior					
County-Level Spying Density	-0.087**	-0.123***	-0.154***	-0.169***	
	(0.034)	(0.039)	(0.049)	(0.049)	
District-Level Spying Density					-0.152**
					(0.062)
Number of Observations	2,946	1,761	1,761	1,761	1,761
Adjusted R-Squared	0.015	0.025	0.106	0.184	0.179
Panel C – Attend Elections					
County-Level Spying Density	-0.035	-0.076	-0.081**	-0.060*	
	(0.037)	(0.058)	(0.031)	(0.031)	
District-Level Spying Density					-0.115***
					(0.039)
Number of Observations	3,041	1,835	1,835	1,835	1,835
Adjusted R-Squared	0.006	0.018	0.128	0.151	0.152
Panel D – Far-Right/Left Orientation					
County-Level Spying Density	0.030	0.033	0.053	0.110**	
	(0.050)	(0.064)	(0.061)	(0.051)	
District-Level Spying Density					0.177**
					(0.069)
Number of Observations	2,474	1,506	1,506	1,506	1,506
Adjusted R-Squared	0.011	0.018	0.104	0.142	0.144
Border County-Pair Fixed Effects			Yes	Yes	Yes
Controls				Yes	Yes

Notes: This table shows the effect of state surveillance on different measures of interpersonal and institutional trust (see panels). The underlying econometric model is described in equation (2.1). In columns (1)–(4), the main regressor of interest is the county-level spying density. In column (5), we combine the border design with an instrumental variables strategy. Here, the leave-out district-level spying density serves as the regressor of interest. All variables have been standardized. All specifications include dummies for historical provinces of the Weimar Republic, year fixed effects, and a dummy variable indicating the presence of a Stasi on-site office. As indicated at the bottom of the table, border county-pair fixed effects and control variables are included in some specifications. These control variables comprise the individuals' age and gender, as well as measures of the size and demographic/industrial composition of the counties in the 1980s, measures for the strength of the opposition to the regime in the 1950s, and indicators for the economic, political, and religious composition of the counties pre-World War II (see Section 2.4.2 for details). While specification (1) is estimated on the full sample of counties, estimates in columns (2)–(5) are based on the sample of contiguous county pairs that straddle a GDR district border. Cross-sectional weights are adjusted for duplicates of counties that are part of multiple pairs. Standard errors are two-way clustered at the county-pair and the county level. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. See Data Appendix 2.B for detailed information on all variables.

standard deviation. Furthermore, we find that a higher spying density induces voters to endorse more politically extreme parties, i.e., to move to the far right or far left of the political spectrum conditional on voting.[24]

Table 2.3 summarizes the results for our measures of economic performance. In line with the trust literature, we also find negative economic effects of Stasi spying. A one standard deviation increase in the spying intensity raises the individual unemployment duration by 0.07-0.11 of a standard deviation and lowers labor income (conditional on working) by 0.11-0.16 of a standard deviation. In contrast, we find small and imprecise effects on the probability to become self-employment (Panel C).

Our estimates reported in Tables 2.2 and 2.3 are standardized to compare the effect size of our estimates across measures, and hence test the internal validity of our results. Reassuringly, estimated effects within types of trust and between measures of trust and economic performance are of similar magnitudes. To ease the interpretation of the effects, we further transform our estimates into a more intuitive scale in terms of outcome variables. Our results imply that a one standard deviation increase in the spying density decreases the probability to trust by 2-3 percentage points and lowers the likelihood of reciprocal behavior by about 2 percentage points. Looking at institutional trust, a one standard deviation higher intensity of surveillance decreases the probability to attend elections by around 2-3 percentage points, while the likelihood to vote for extreme parties increases by 3-5 percentage points. As regards economic measures, the individual unemployment exposure increases by 1-2 percentage points, while monthly gross income – conditional on working – is 100-140 EUR lower.

2.5.2 Sensitivity Checks and Additional Results

We next provide a range of robustness checks to make sure that estimates are not driven by modeling choices.

Inference. Standard errors of our baseline results are two-way clustered at the county-pair and county level. As discussed above, we choose this default to account for common shocks within county pairs as well as the duplication of certain counties. One-way clustering at the county-pair level yields very similar standard errors. However, parts of the identifying variation are induced by differences in surveillance strategies across districts, which would imply that standard errors should be clustered at the district level. This might, in turn, be problematic due to the small number of clusters ($N = 14$). We account for this by implementing the standard percentile-t Wild cluster bootstrap approach as proposed by Cameron et al. (2008). Panel B in

[24] Results are equally driven by moves towards both sides of the political spectrum.

Table 2.3: The Effect of Spying on Economic Performance

	All Counties	Border County Sample			
	(1)	(2)	(3)	(4)	(5)
Panel A – Unemployment Duration					
County-Level Spying Density	0.026	0.059	0.054	0.072**	
	(0.028)	(0.048)	(0.035)	(0.030)	
District-Level Spying Density					0.113***
					(0.031)
Number of Observations	2,795	1,739	1,739	1,739	1,739
Adjusted R-Squared	0.007	0.018	0.091	0.131	0.132
Panel B – Log Mean Labor Income					
County-Level Spying Density	-0.067**	-0.072*	-0.085**	-0.155***	
	(0.033)	(0.043)	(0.035)	(0.037)	
District-Level Spying Density					-0.113**
					(0.056)
Number of Observations	2,308	1,422	1,422	1,422	1,422
Adjusted R-Squared	0.014	0.013	0.069	0.171	0.165
Panel C – Self-Employment Duration					
County-Level Spying Density	0.053**	0.039	-0.005	0.001	
	(0.024)	(0.034)	(0.032)	(0.019)	
District-Level Spying Density					-0.004
					(0.026)
Number of Observations	2,792	1,739	1,739	1,739	1,739
Adjusted R-Squared	0.007	0.009	0.046	0.070	0.070
Border County-Pair Fixed Effects			Yes	Yes	Yes
Controls				Yes	Yes

Notes: This table shows the effect of state surveillance on different measures of economic performance (see panels). The underlying econometric model is described in equation (2.1). In columns (1)–(4), the main regressor of interest is the county-level spying density. In column (5), we combine the border design with an instrumental variables strategy. Here, the leave-out district-level spying density serves as the regressor of interest. All variables have been standardized. All specifications include dummies for historical provinces of the Weimar Republic, year fixed effects, and a dummy variable indicating the presence of a Stasi on-site office. As indicated at the bottom of the table, border county-pair fixed effects and control variables are included in some specifications. These control variables comprise the individuals' age and gender, as well as measures of the size and demographic/industrial composition of the counties in the 1980s, measures for the strength of the opposition to the regime in the 1950s, and indicators for the economic, political, and religious composition of the counties pre-World War II (see Section 2.4.2 for details). While specification (1) is estimated on the full sample of counties, estimates in columns (2)–(5) are based on the sample of contiguous county pairs that straddle a GDR district border. Cross-sectional weights are adjusted for duplicates of counties that are part of multiple pairs. Standard errors are two-way clustered at the county-pair and the county level. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. See Data Appendix 2.B for detailed information on all variables.

Appendix Table 2.A.2 shows that reduced form estimates are at least statistically significant at the 10 percent level – with the exception of labor income, which is significant at the 13 %-level.

As an alternative robustness check, we further conduct randomization inference in the spirit of Chetty et al. (2009) and Fouka and Voth (2016). We perform 2,999 random permutations of the dependent variable and re-estimate model (2.1) each time. The corresponding p-values reported in Table 2.A.2 indicate the share of estimated coefficients that are smaller (larger for the outcomes "far right/left orientation" and "unemployment duration") than our estimates. Again, results suggests that our inference is robust.

Deep Historical Differences. Tables 2.2 and 2.3 show that accounting for observable historical differences within county pairs increases the magnitude of the estimates. In addition, we can exploit an institutional feature of the GDR to further improve the comparability of counties within pairs. To this end, we draw on the territorial reforms of 1949–1952 that eliminated the decentralized structure of the Weimar Republic and created 15 purely administrative districts (cf. Section 2.2). The explicit goal of the reform was to remove the influence of the formerly powerful provinces of the Weimar Republic. As a consequence, new borders were often drawn through former Weimar provinces, separating regions with the same cultural heritage (see Figure 2.A.4). If unobserved cultural differences within county pairs at newly drawn borders are smaller than differences at historical borders, potential endogeneity bias should be smaller for the former type of pairs. We assess this hypothesis by estimating an interaction model differentiating between historical and newly drawn borders.[25] We generally find larger absolute effects and more precise estimates for county pairs which share the same cultural heritage (see Table 2.A.3). We take this as further evidence that endogeneity induces an upward bias and estimates become larger in absolute terms as we tighten the identification strategy.

County Pairs and Weighting. As discussed in Section 2.4.1, there are various ways to define the estimation sample in case of multiple neighboring counties. To limit the number of duplications and find pairs that are largely similar in observable characteristics, we only consider the neighbor that shares the longest borderline with the respective county in our baseline specification. Nevertheless, we also provide estimates based on the full set of county pairs. As a third variant, we consider a specification without any duplicates, discarding relatively smaller pairs that would lead to a duplication of counties. Results are summarized in Appendix

[25] Precisely, we interact the spying density of our baseline border design model given in equation (2.1) with a dummy variable indicating whether counties within a pair are from the same or from different Weimar provinces. We also exclude Weimar province dummies from the regressions to be able to also identify effects of differences between former provinces. Dropping these province dummies considerably lowers the power of the instrument, so we do not report reduced form estimates for this specification.

Table 2.A.4.

Overall, we find that results from our baseline model are on average larger in absolute terms compared to estimates obtained from the full set of possible county pairs. This pattern is in line with the results from the smoothness test shown in Table 2.1, which indicate that our baseline specification does better in balancing covariates, thereby lending more credibility to the identifying assumption of the border design. Results also prevail when avoiding any duplication despite the much smaller sample. Last, Appendix Table 2.A.4 shows that estimates are also robust with respect to the duplication adjustment of sample weights.

Other Measures of Surveillance Intensity and Alternative Channels. While the number of operative informers is arguably the most natural measure of surveillance intensity (cf. Section 2.2), we show in Appendix Table 2.A.5 that results are robust when using alternative definitions such as the total number of informers or when additionally including the number of official Stasi employees. Finally, we show that our estimates are robust to controlling for regional variation in political indoctrination.

2.6 Conclusion

In this paper, we investigate the effect of state surveillance on interpersonal and institutional trust and economic performance. We study the case of the former socialist German Democratic Republic, one of the largest surveillance systems of all time, and exploit county-level variation in the density of Stasi informers. To account for the non-random recruitment of informers across counties, we harness the specific institutional features of the East German surveillance state and combine a standard border discontinuity design with an instrumental variables approach.

Overall, the results of our study offer substantial evidence for negative and long-lasting effects of government surveillance. We find strong and consistent evidence that a higher density of informers undermines individuals' interpersonal and institutional trust. In particular, more intense surveillance causes lower trust in strangers, less positive reciprocity, lower participation in elections, and more support for extreme political parties. Against this backdrop, we further find negative and persistent effects of government surveillance on measures of economic performance, as measured by individual labor income and unemployment duration.

Our results add to the literature on institutions, trust, and economic performance (Alesina and Giuliano, 2015). First, our study establishes a causal link between formal institutions (surveillance) and culture (trust). Second, and in line with Tabellini (2010), we provide evidence that the degree of democratic governance affects economic outcomes. Third, with both trust and economic performance being impaired by government surveillance, our findings also provide suggestive evidence in favor of a well-established channel: institutions shape people's trust, and trust affects economic development (Algan and Cahuc, 2014). In this respect, government surveillance might have had an effect on economic performance through alternative channels, such as personality traits or liberal values. While investigating these channels is certainly interesting, we follow the literature and focus on the broader outcome of trust, which is predicted to affect economic performance (Berg et al., 1995) and has been shown to be shaped by personality traits (Dohmen et al., 2008). Last, we show that our effects are persistent and still detectable two decades after the end of the socialist regime. Given the intergenerational transmission of trust and beliefs (see, e.g., Nunn and Wantchekon, 2011, Dohmen et al., 2012), it is possible that these effects will be even longer-lasting.

One important question is how our findings translate to other (contemporary) forms of mass surveillance in authoritarian states given that surveillance strategies have changed over the last decades and nowadays rely arguably more on technology than individual informers.[26] It is likely that this shift towards electronic surveillance modes renders the findings for interpersonal trust within the social network less important. At the same time, it seems plausible that trust in institutions could still be affected by modern forms of surveillance. After the revelation of the NSA wiretapping and the Snowden affair, for example, anecdotal evidence suggests that citizens did not know which communication companies to trust (see, e.g., Schneier, 2013). Moreover, a large share of people stated that they had adjusted their use of telecommunications as a consequence of the affair (Pew Research Center, 2014).

The Snowden affair further points to another conceptual issue when generalizing our findings. Are effects of government surveillance different in a democracy? Both democratic and autocratic regimes would justify surveillance with the need to secure the stability of the system – hence with benevolent motives, while the (perceived) degree of benevolence is, of course, highly subjective. Separating negative and positive aspects of surveillance is notoriously difficult, and researchers will most likely only be able to assess the net effect of surveillance. The findings of this study show that the net effect of government surveillance on trust and economic performance was negative in the case of socialist East Germany. Net effects of state surveillance in other systems and at different times may vary and should to be studied case-by-case.

[26] Nevertheless, contemporaneous regimes still make use of informers to control their citizens. Various accounts state that China still heavily relies on a large network of informers (see, e.g., Branigan, 2010, Jacobs and Ansfield, 2011, Yu, 2014). Likewise, Russia has been observed to re-implement surveillance strategies in which secret informers and denunciations play an important role in controlling opposition forces (Capon, 2015).

Appendix 2.A Further Results and Figures

Table 2.A.1: The Effect of Spying on Alternative Measures of Trust

	All Counties	Border County Sample			
	(1)	(2)	(3)	(4)	(5)
Panel A – Number of Friends					
County-Level Spying Density	-0.045	-0.054	-0.084**	-0.089**	
	(0.051)	(0.036)	(0.037)	(0.037)	
District-Level Spying Density					-0.074*
					(0.043)
Number of Observations	3,110	1,903	1,903	1,903	1,903
Adjusted R-Squared	0.012	0.026	0.105	0.123	0.121
Panel B – Engagement in Local Politics					
County-Level Spying Density	0.032	-0.005	-0.071**	-0.098***	
	(0.025)	(0.037)	(0.033)	(0.028)	
District-Level Spying Density					-0.114***
					(0.039)
Number of Observations	3,555	2,166	2,166	2,166	2,166
Adjusted R-Squared	0.006	0.013	0.054	0.071	0.070
Border County-Pair Fixed Effects			Yes	Yes	Yes
Controls				Yes	Yes

Notes: This table shows the effect of state surveillance on alternative measures of interpersonal and institutional trust (see panels). Results for the baseline measures of trust are shown in Table 2.2. The underlying econometric model is described in equation (2.1). In columns (1)–(4), the main regressor of interest is the county-level spying density. In column (5), we combine the border design with an instrumental variables strategy. Here, the leave-out district-level spying density serves as the regressor of interest. All variables have been standardized. All specifications include dummies for historical provinces of the Weimar Republic, year fixed effects, and a dummy variable indicating the presence of a Stasi on-site office. As indicated at the bottom of the table, border county-pair fixed effects and control variables are included in some specifications. These control variables comprise the individuals' age and gender, as well as measures of the size and demographic/industrial composition of the counties in the 1980s, measures for the strength of the opposition to the regime in the 1950s, and indicators for the economic, political, and religious composition of the counties pre-World War II (see Section 2.4.2 for details). While specification (1) is estimated on the full sample of counties, estimates in columns (2)–(5) are based on the sample of contiguous county pairs that straddle a GDR district border. Cross-sectional weights are adjusted for duplicates of counties that are part of multiple pairs. Standard errors are two-way clustered at the county-pair and the county level. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. See Data Appendix 2.B for detailed information on all variables.

Table 2.A.2: Robustness Checks – Inference

	Trust in Strangers (1)	Reciprocal Behavior (2)	Attend Elections (3)	Extreme Left/Right (4)	Unemploy. Duration (5)	Labor Income (6)	Self-Emp. Duration (7)
Panel A – County-Level Spying Density							
Baseline Estimate	-0.065	-0.169	-0.060	0.110	0.072	-0.155	0.001
Cluster on County-Pair and County Level	(0.027)	(0.049)	(0.031)	(0.051)	(0.030)	(0.037)	(0.019)
	[0.022]	[0.001]	[0.055]	[0.034]	[0.020]	[0.000]	[0.976]
Alternative Cluster Definitions							
Cluster on County-Pair Level	(0.028)	(0.047)	(0.031)	(0.050)	(0.034)	(0.037)	(0.020)
	[0.023]	[0.001]	[0.055]	[0.032]	[0.036]	[0.000]	[0.977]
Cluster on County Level	(0.021)	(0.039)	(0.024)	(0.038)	(0.022)	(0.028)	(0.015)
	[0.003]	[0.000]	[0.015]	[0.005]	[0.001]	[0.000]	[0.969]
Cluster on County-Pair and District Level	(0.025)	(0.050)	(0.032)	(0.040)	(0.030)	(0.035)	(0.020)
	[0.023]	[0.005]	[0.080]	[0.017]	[0.031]	[0.001]	[0.978]
Wild Cluster Bootstrap-t (H_0 imposed)							
Cluster on County-Pair and District Level	[0.090]	[0.047]	[0.161]	[0.059]	[0.172]	[0.019]	[0.983]
Randomization Inference							
Cumulative Distribution of Estimates	[0.082]	[0.000]	[0.095]	[0.023]	[0.038]	[0.000]	[0.522]
Panel B – District-Level Spying Density							
Baseline Estimate	-0.092	-0.152	-0.115	0.177	0.113	-0.113	-0.004
Cluster on County-Pair and County Level	(0.044)	(0.062)	(0.039)	(0.069)	(0.031)	(0.056)	(0.026)
	[0.042]	[0.019]	[0.004]	[0.014]	[0.001]	[0.049]	[0.871]
Alternative Cluster Definitions							
Cluster on County-Pair Level	(0.043)	(0.064)	(0.040)	(0.068)	(0.035)	(0.054)	(0.027)
	[0.040]	[0.021]	[0.006]	[0.012]	[0.002]	[0.043]	[0.877]
Cluster on County Level	(0.036)	(0.049)	(0.031)	(0.054)	(0.022)	(0.043)	(0.020)
	[0.013]	[0.003]	[0.000]	[0.002]	[0.000]	[0.011]	[0.830]
Cluster on County-Pair and District Level	(0.035)	(0.068)	(0.041)	(0.069)	(0.027)	(0.052)	(0.027)
	[0.022]	[0.045]	[0.015]	[0.025]	[0.001]	[0.051]	[0.880]
Wild Cluster Bootstrap-t (H_0 imposed)							
Cluster on County-Pair and District Level	[0.080]	[0.082]	[0.055]	[0.041]	[0.011]	[0.134]	[0.919]
Randomization Inference							
Cumulative Distribution of Estimates	[0.051]	[0.007]	[0.029]	[0.005]	[0.013]	[0.031]	[0.488]

Notes: This table presents robustness checks on inference for our baseline estimates based on the county-level spying density (cf. column (4) of Tables 2.2 and 2.3) in Panel A and our reduced form estimate using the leave-out spying density at the district level (cf. column (5) of Tables 2.2 and 2.3) in Panel B. On top of each panel, we present the corresponding point estimate along with our baseline standard errors that are two-way clustered at the county-pair and the county level (in parentheses) and the corresponding p-values in square brackets. Below, we report the corresponding standard errors and p-values for alternative cluster specifications, as well as empirical p-values from a wild cluster percentile-t bootstrap test with H_0 imposed (Cameron et al., 2008). In the last row of each panel, we present p-values based on randomization inference in the spirit of Chetty et al. (2009) and Fouka and Voth (2016). Empirical p-values from the wild cluster percentile-t bootstrap and randomization inference procedures are both based on 2,999 replications. All specifications include border county-pair fixed effects, dummies for historical provinces of the Weimar Republic, year fixed effects, a dummy variable indicating the presence of a Stasi on-site office as well as control variables which comprise the individuals' age and gender, as well as measures of the size and demographic/industrial composition of the counties in the 1980s, measures for the strength of the opposition to the regime in the 1950s, and indicators for the economic, political, and religious composition of the counties pre-World War II (see Section 2.4.2 for details). All specifications are based on the sample of contiguous county pairs that straddle a GDR district border. Cross-sectional weights are adjusted for duplicates of counties that are part of multiple pairs. See Data Appendix 2.B for detailed information on all variables.

Table 2.A.3: The Effect of Spying by Weimar Provinces

	Trust in Strangers (1)	Reciprocal Behavior (2)	Attend Elections (3)	Extreme Left/Right (4)	Unemploy. Duration (5)	Labor Income (6)	Self-Emp. Duration (7)
County-Level Spying Density							
× Different Weimar Province	-0.065	0.058	-0.054	0.041	-0.023	0.058	-0.031
	(0.048)	(0.089)	(0.067)	(0.091)	(0.083)	(0.107)	(0.035)
× Same Weimar Province	-0.075**	-0.253***	-0.050	0.089	0.097**	-0.186***	0.041
	(0.034)	(0.085)	(0.039)	(0.069)	(0.044)	(0.053)	(0.027)
Number of Observations	2,007	1,761	1,835	1,506	1,739	1,422	1,739

Notes: This table shows the effect of state surveillance on trust and economic performance when allowing for heterogeneous treatment effects for county pairs with similar and different cultural heritage. The underlying econometric model is based on equation (2.1) where the county-level spying density is fully interacted with a dummy variable indicating whether two counties in a pair belonged to the same state or Prussian province of the Weimar Republic before World War II. All variables have been standardized. All specifications include border county-pair fixed effects, year fixed effects, a dummy variable indicating the presence of a Stasi on-site office as well as control variables which comprise the individuals' age and gender, as well as measures of the size and demographic/industrial composition of the counties in the 1980s, measures for the strength of the opposition to the regime in the 1950s, and indicators for the economic, political, and religious composition of the counties pre-World War II (see Section 2.4.2 for details). All regressions are based on the sample of contiguous county pairs that straddle a GDR district border. Cross-sectional weights are adjusted for duplicates of counties that are part of multiple pairs. Standard errors are two-way clustered at the county-pair and the county level. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. See Data Appendix 2.B for detailed information on all variables.

Table 2.A.4: Robustness Checks – County-Pair Sample Definition

	Baseline		All Pairs		No Duplicates		No Wgt. Adjust.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A – Trust in Strangers								
County-Level Spying Density	-0.065**		-0.081**		-0.096**		-0.064**	
	(0.027)		(0.033)		(0.041)		(0.031)	
District-Level Spying Density		-0.092**		-0.014		-0.237***		-0.071
		(0.044)		(0.046)		(0.051)		(0.045)
Number of Observations	2,007	2,007	2,814	2,814	1,226	1,226	2,007	2,007
Panel B – Reciprocal Behavior								
County-Level Spying Density	-0.169***		-0.142***		-0.230***		-0.130**	
	(0.049)		(0.040)		(0.053)		(0.050)	
District-Level Spying Density		-0.152**		-0.114**		-0.232***		-0.117*
		(0.062)		(0.056)		(0.068)		(0.060)
Number of Observations	1,761	1,761	2,451	2,451	1,075	1,075	1,761	1,761
Panel C – Attend Elections								
County-Level Spying Density	-0.060*		-0.089***		-0.095**		-0.055*	
	(0.031)		(0.031)		(0.037)		(0.032)	
District-Level Spying Density		-0.115***		-0.064*		-0.091**		-0.115***
		(0.039)		(0.034)		(0.040)		(0.038)
Number of Observations	1,835	1,835	2,569	2,569	1,127	1,127	1,835	1,835
Panel D – Far-Right/Left Orientation								
County-Level Spying Density	0.110**		0.080*		0.059		0.104**	
	(0.051)		(0.041)		(0.075)		(0.048)	
District-Level Spying Density		0.177**		0.101**		0.074		0.154**
		(0.069)		(0.048)		(0.113)		(0.059)
Number of Observations	1,506	1,506	2,111	2,111	910	910	1,506	1,506
Panel E – Unemployment Duration								
County-Level Spying Density	0.072**		0.015		0.066*		0.069**	
	(0.030)		(0.035)		(0.038)		(0.034)	
District-Level Spying Density		0.113***		0.065		0.053*		0.121***
		(0.031)		(0.042)		(0.031)		(0.029)
Number of Observations	1,739	1,739	2,366	2,366	1,036	1,036	1,739	1,739
Panel F – Log Mean Labor Income								
County-Level Spying Density	-0.155***		-0.114***		-0.231***		-0.147***	
	(0.037)		(0.041)		(0.059)		(0.037)	
District-Level Spying Density		-0.113**		-0.065		-0.209***		-0.091*
		(0.056)		(0.048)		(0.074)		(0.048)
Number of Observations	1,422	1,422	1,943	1,943	839	839	1,422	1,422
Panel G – Self-Employment Duration								
County-Level Spying Density	0.001		-0.008		0.044*		-0.001	
	(0.019)		(0.019)		(0.023)		(0.018)	
District-Level Spying Density		-0.004		-0.027		0.008		-0.017
		(0.026)		(0.028)		(0.028)		(0.024)
Number of Observations	1,739	1,739	2,366	2,366	1,036	1,036	1,739	1,739

Notes: This table shows the effects of state surveillance on trust and economic performance when using different definitions of the county-pair sample. The underlying econometric model is described in equation (2.1). All variables have been standardized. Estimates in columns (1)–(2) print our baseline results shown in columns (4) and (5) of Tables 2.2 and 2.3. These estimates are based on our preferred county-pair sample and adjusted weights for the duplication of counties. Columns (3) and (4) show the corresponding results when using all existing county pairs. In turn, estimates in columns (5)–(6) are based on a smaller sample that excludes all county duplicates. Estimates in columns (7) and (8) are based on our baseline definition of the county-pair sample but weights are not adjusted if counties are duplicated. All specifications use cross-sectional weights and include border county-pair fixed effects, dummies for historical provinces of the Weimar Republic, year fixed effects, a dummy variable indicating the presence of a Stasi on-site office as well as control variables which comprise the individuals' age and gender, as well as measures of the size and demographic/industrial composition of the counties in the 1980s, measures for the strength of the opposition to the regime in the 1950s, and indicators for the economic, political, and religious composition of the counties pre-World War II (see Section 2.4.2 for details). Standard errors are two-way clustered at the county-pair and the county level. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. See Data Appendix 2.B for detailed information on all variables.

Table 2.A.5: Robustness Checks – Alternative Measures

	Baseline		IM1 and IM2		IM1, IM2, HAM		Control f. Indoctr.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A – Trust in Strangers								
County-Level Spying Density	-0.065**		-0.065**		-0.068**		-0.056**	
	(0.027)		(0.027)		(0.027)		(0.023)	
District-Level Spying Density		-0.092**		-0.120**		-0.119**		-0.084**
		(0.044)		(0.046)		(0.055)		(0.037)
Number of Observations	2,007	2,007	1,743	1,743	1,743	1,743	2,007	2,007
Panel B – Reciprocal Behavior								
County-Level Spying Density	-0.169***		-0.147***		-0.153***		-0.173***	
	(0.049)		(0.052)		(0.051)		(0.052)	
District-Level Spying Density		-0.152**		-0.195***		-0.191**		-0.154**
		(0.062)		(0.067)		(0.072)		(0.063)
Number of Observations	1,761	1,761	1,528	1,528	1,528	1,528	1,761	1,761
Panel C – Attend Elections								
County-Level Spying Density	-0.060*		-0.093***		-0.098***		-0.059*	
	(0.031)		(0.030)		(0.029)		(0.030)	
District-Level Spying Density		-0.115***		-0.146***		-0.140***		-0.114***
		(0.039)		(0.041)		(0.043)		(0.037)
Number of Observations	1,835	1,835	1,595	1,595	1,595	1,595	1,835	1,835
Panel D – Far-Right/Left Orientation								
County-Level Spying Density	0.110**		0.095**		0.098**		0.118**	
	(0.051)		(0.047)		(0.047)		(0.053)	
District-Level Spying Density		0.177**		0.191**		0.213**		0.180**
		(0.069)		(0.076)		(0.090)		(0.070)
Number of Observations	1,506	1,506	1,321	1,321	1,321	1,321	1,506	1,506
Panel E – Unemployment Duration								
County-Level Spying Density	0.072**		0.065*		0.063*		0.067**	
	(0.030)		(0.033)		(0.034)		(0.030)	
District-Level Spying Density		0.113***		0.159***		0.158***		0.109***
		(0.031)		(0.036)		(0.041)		(0.032)
Number of Observations	1,739	1,739	1,526	1,526	1,526	1,526	1,739	1,739
Panel F – Log Mean Labor Income								
County-Level Spying Density	-0.155***		-0.156***		-0.157***		-0.150***	
	(0.037)		(0.042)		(0.042)		(0.037)	
District-Level Spying Density		-0.113**		-0.135*		-0.110		-0.112*
		(0.056)		(0.068)		(0.080)		(0.058)
Number of Observations	1,422	1,422	1,245	1,245	1,245	1,245	1,422	1,422
Panel G – Self-Employment Duration								
County-Level Spying Density	0.001		0.015		0.012		0.000	
	(0.019)		(0.018)		(0.019)		(0.019)	
District-Level Spying Density		-0.004		0.010		0.003		-0.005
		(0.026)		(0.031)		(0.034)		(0.025)
Number of Observations	1,739	1,739	1,526	1,526	1,526	1,526	1,739	1,739

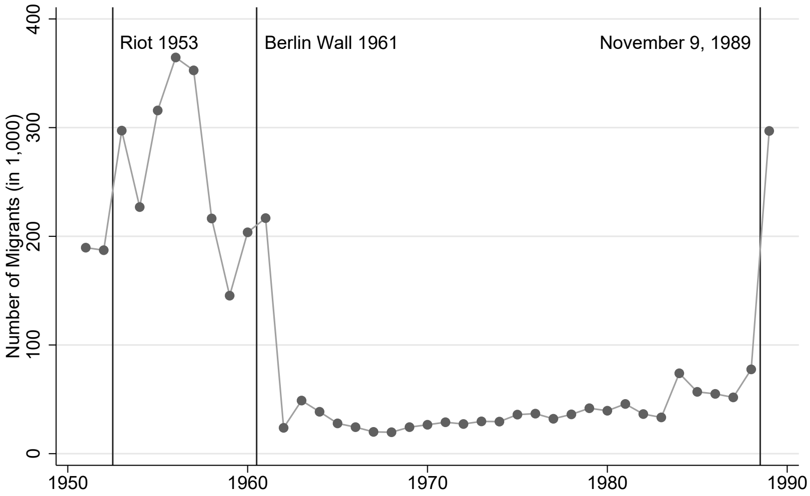
Notes: This table shows the effects of state surveillance on trust and economic performance when using different measures of government surveillance and accounting for alternative channels. The underlying econometric model is described in equation (2.1). All variables have been standardized. Estimates in columns (1)–(2) print our baseline results shown in columns (4) and (5) of Tables 2.2 and 2.3. These estimates use the share of operative informers (*IM1*) in the population as our measure of state surveillance. Columns (3) and (4) show the corresponding results when taking the per-capita number of operative informers (*IM1*) and informers providing logistics (*IM2*) as our measure of state surveillance. In columns (5)–(6), the measure of spying density includes operative informers (*IM1*), informers providing logistics (*IM2*), and official Stasi employees. Columns (7) and (8) show the results for our baseline specification when additionally controlling for differences in the political indoctrination of counties (measured by the share of SED party members among the local elites). All specifications include border county-pair fixed effects, dummies for historical provinces of the Weimar Republic, year fixed effects, a dummy variable indicating the presence of a Stasi on-site office as well as control variables which comprise the individuals' age and gender, as well as measures of the size and demographic/industrial composition of the counties in the 1980s, measures for the strength of the opposition to the regime in the 1950s, and indicators for the economic, political, and religious composition of the counties pre-World War II (see Section 2.4.2 for details). All specifications are based on the sample of contiguous county pairs that straddle a GDR district border. Cross-sectional weights are adjusted for duplicates of counties that are part of multiple pairs. Standard errors are two-way clustered at the county-pair and the county level. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. See Data Appendix 2.B for detailed information on all variables.

Table 2.A.6: Instrumental Variables Results

	Leave-Out District Mean			District-Level Average		
	First Stage (1)	Reduced Form (2)	Second Stage (3)	First Stage (4)	Reduced Form (5)	Second Stage (6)
Panel A – Trust in Strangers (N = 2, 007)						
District-Level Spying Density	0.777***	-0.092**		0.885***	-0.084**	
	(0.231)	(0.044)		(0.156)	(0.035)	
County-Level Spying Density			-0.118**			-0.095**
			(0.051)			(0.038)
Weak Instrument F-Statistic	11.35			32.32		
Underidentification p-Value	0.024			0.010		
Panel B – Reciprocal Behavior (N = 1, 761)						
District-Level Spying Density	0.816***	-0.152**		0.893***	-0.169***	
	(0.212)	(0.062)		(0.144)	(0.059)	
County-Level Spying Density			-0.186**			-0.189***
			(0.085)			(0.069)
Weak Instrument F-Statistic	14.86			38.23		
Underidentification p-Value	0.016			0.008		
Panel C – Attend Elections (N = 1, 835)						
District-Level Spying Density	0.821***	-0.115***		0.900***	-0.107***	
	(0.214)	(0.039)		(0.143)	(0.036)	
County-Level Spying Density			-0.140**			-0.119**
			(0.067)			(0.051)
Weak Instrument F-Statistic	14.73			39.91		
Underidentification p-Value	0.020			0.010		
Panel D – Far-Right/Left Orientation (N = 1, 506)						
District-Level Spying Density	0.763***	0.177**		0.863***	0.175**	
	(0.222)	(0.069)		(0.153)	(0.067)	
County-Level Spying Density			0.231*			0.203**
			(0.125)			(0.094)
Weak Instrument F-Statistic	11.85			31.61		
Underidentification p-Value	0.020			0.009		
Panel E – Unemployment Duration (N = 1, 739)						
District-Level Spying Density	0.683***	0.113***		0.789***	0.107***	
	(0.212)	(0.031)		(0.155)	(0.029)	
County-Level Spying Density			0.166***			0.135***
			(0.058)			(0.039)
Weak Instrument F-Statistic	10.33			26.01		
Underidentification p-Value	0.026			0.011		
Panel F – Log Mean Labor Income (N = 1, 422)						
District-Level Spying Density	0.698***	-0.113**		0.788***	-0.122**	
	(0.195)	(0.056)		(0.145)	(0.051)	
County-Level Spying Density			-0.161***			-0.154***
			(0.054)			(0.046)
Weak Instrument F-Statistic	12.78			29.72		
Underidentification p-Value	0.020			0.010		
Panel G – Self-Employment Duration (N = 1, 739)						
District-Level Spying Density	0.683***	-0.004		0.789***	-0.003	
	(0.212)	(0.026)		(0.155)	(0.023)	
County-Level Spying Density			-0.006			-0.004
			(0.038)			(0.029)
Weak Instrument F-Statistic	10.33			26.01		
Underidentification p-Value	0.026			0.011		

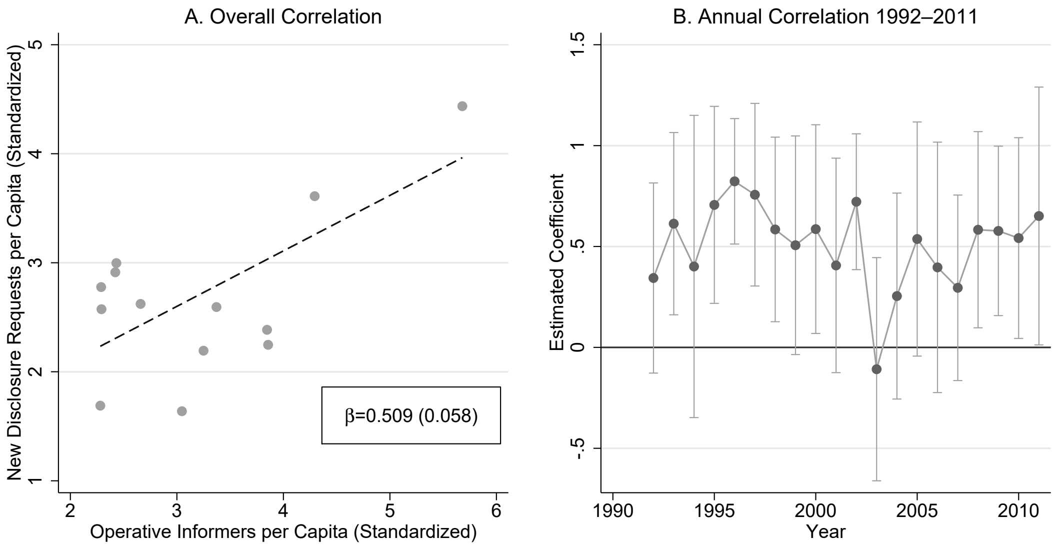
Notes: This table provides more detailed results on the effects of state surveillance on trust and economic performance when combining the border design with an instrumental variables strategy. The underlying econometric model is described in equation (2.1), where we either use the leave-out spying density at the district level as the instrument (columns (1)–(3)) or the simple district average (columns (4)–(6)). All variables have been standardized. Estimates in column (2) print our baseline reduced form results shown in column (5) of Tables 2.2 and 2.3. Columns (1) and (4) report first-stage results as well as test for the strength and relevance of the respective instruments. Columns (3) and (6) report the corresponding second-stage results. All specifications include border county-pair fixed effects, dummies for historical provinces of the Weimar Republic, year fixed effects, a dummy variable indicating the presence of a Stasi on-site office as well as control variables which comprise the individuals' age and gender, as well as measures of the size and demographic/industrial composition of the counties in the 1980s, measures for the strength of the opposition to the regime in the 1950s, and indicators for the economic, political, and religious composition of the counties pre-World War II (see Section 2.4.2 for details). All specifications are based on the sample of contiguous county pairs that straddle a GDR district border. Cross-sectional weights are adjusted for duplicates of counties that are part of multiple pairs. Standard errors are two-way clustered at the county-pair and the county level. Significance levels are * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. See Data Appendix 2.B for detailed information on all variables.

Figure 2.A.1: Out-Migration from East Germany



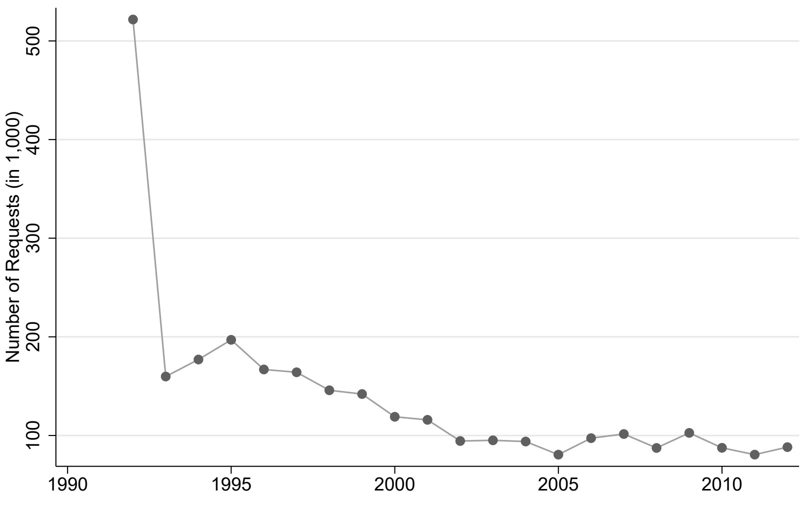
Notes: Based on data from Statistisches Bundesamt (1993).

Figure 2.A.2: Regional Disclosure Requests and Number of Informers



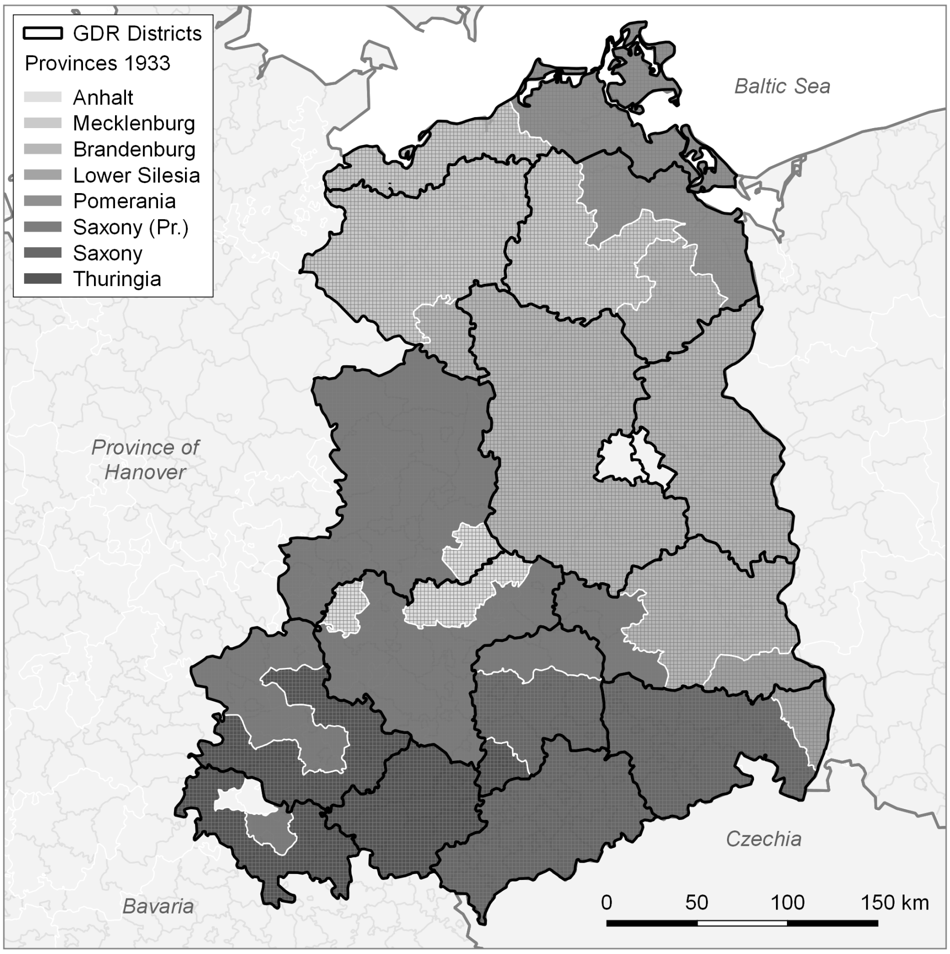
Notes: This graph plots the correlation between the regional number of disclosure requests and the number of operative Stasi informers at the level of the 15 GDR districts, relative to a district's average population from 1980–1988. Panel A shows the overall correlation, controlling for year fixed effects. Panel B shows the evolution of this relationship over time, regressing the number of disclosure requests per capita on the share of operative informers in the 1980s separately year-by-year. Both measures are standardized. Standard errors are heteroscedasticity-robust, Panel B shows 95 % confidence intervals. Data on the district-level number of requests has been provided by the Stasi Records Agency in response to a Freedom of Information request.

Figure 2.A.3: Number of Requests for the Inspection of Stasi Files



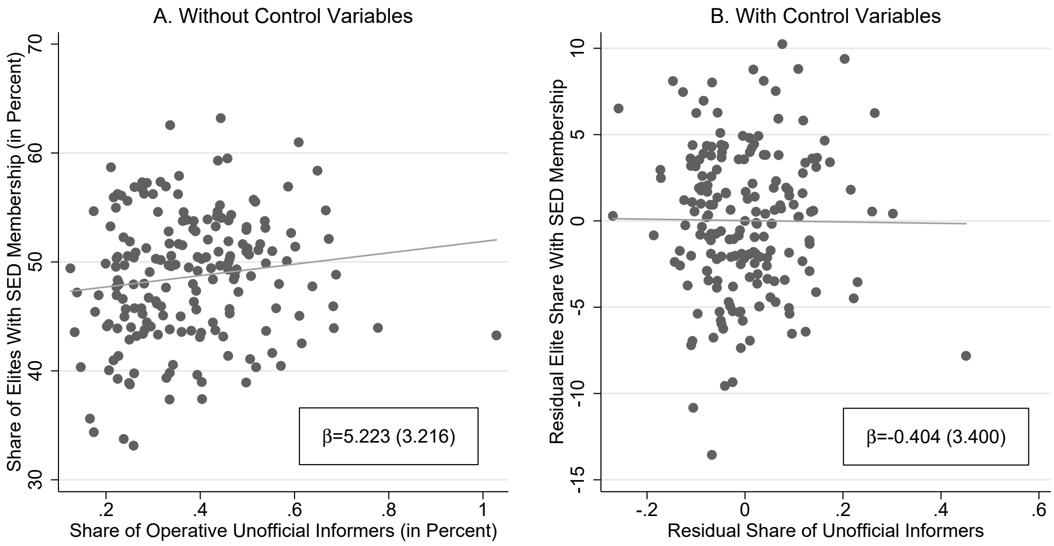
Notes: This graph plots the annual number of requests to inspect Stasi files. It is based on data from the Agency of the Federal Commissioner for the Stasi Records.

Figure 2.A.4: GDR Districts and Provinces of the Weimar Republic



Notes: The figure shows GDR district borders and historical borders of the states of the Weimar Republic and the Prussian provinces as of 1933. *Maps:* MPIDR and CGG (2011) and EuroGeographics.

Figure 2.A.5: Informer Density and Socialist Indoctrination



Notes: The graph plots the correlation between the share of operative unofficial informers and the intensity of socialist indoctrination as measured by the share of elites with a SED party membership. Panel A depicts the unconditional correlation, Panel B the residual correlation after conditioning on county level covariates. The β coefficients report the regression slope, heteroscedasticity-robust standard errors in parentheses. For information on all variables, see Appendix Table 2.B.1.

Appendix 2.B Data Appendix

2.B.1 Variable Definitions and Descriptive Statistics

Table 2.B.1: Definition of Variables and Data Sources

Variable	Years	Source
Panel A – Stasi Data (See Section 2.2)		
Spying Density	1980–1988	The main explanatory variable of interest, the regional spying density, is calculated as the average spying density at the county level in the period 1980–1988 (see Section 2.2 for details). Data on unofficial informers are based on official Stasi records that were in large part compiled by Müller-Enbergs (2008). As the Stasi Records Agency keeps restoring files and releasing new data, we collected additional information from the Stasi archives and expanded the data set with informer figures from ten previously missing county offices (*Akteneinsicht zu Forschungszwecken, BStU Tgb. 15582/15Z*). Population figures were taken from the Statistical Yearbooks of the GDR. Our baseline measure of spying density covers unofficial informers for political-operative penetration, homeland defense, or special operations, as well as leading informers (*IM zur politisch-operativen Durchdringung und Sicherung des Verantwortungsbereiches, IM der Abwehr mit Feindverbindung bzw. zur unmittelbaren Bearbeitung im Verdacht der Feindtätigkeit stehender Personen, IM im besonderen Einsatz, Führungs-IM*). In cases where the Stasi held additional on-site offices (*Objektdienststellen*) in counties, the number of informers attached to these offices was added to the number of spies in the respective county.
Stasi Employees	1982	The number of regular Stasi employees (*Hauptamtliche Mitarbeiter*) attached to county offices in 1982 was provided by Jens Gieseke.
Panel B – Individual SOEP Data (See Section 2.4.2)		
Attend Elections	2005, 2009	Respondents are asked about their willingness to attend the next election for the German parliament. The question reads as follows: "If the next election to the German 'Bundestag' (lower house of parliament) were next Sunday, would you vote?" Response options were given on a five-point scale allowing individuals to express varying degrees of conviction (not) to vote. We re-scaled the original variable such that higher values indicate a stronger willingness to participate. We discard individuals that are not eligible to vote.

continued

Table 2.B.1 continued

Variable	Years	Source
Far-Right/Left Orientation	2005, 2009	Respondents are asked to state their political orientation, the underlying question being: “In politics, people often talk about ‘left’ and ‘right’ when describing different political views. When you think about your own political view, how would you rate them on the scale below?” Response options were given on a eleven-point scale allowing different placement along the political spectrum. We consider respondents to be political extreme if they are in the upper/lower four percent of the distribution.
Labor Income	1992–2010	Information on current monthly gross labor income is provided in every wave of the SOEP for East German respondents since 1992. We calculate real income in 2010 prices using the official German CPI (*Verbraucherpreisindex*). We discard individuals with labor income below 100 EUR.
Reciprocal Behavior	2005, 2010	We use six questions on positive and negative reciprocity to combine them into one single measure by taking the simple mean. Response options on each statement vary on a seven-point scale and we re-scale the responses on the three statements indicating negative reciprocity such that higher values in our variable indicate more positive reciprocal behavior.
Self-Employment Probability	1992–2009	Detailed information on individuals’ type of employment is given in every wave of the survey. The data set distinguishes between self-employed farmers, free-lance professionals, solo self-employed and self-employed individuals with coworkers. We focus on the latter category and construct our measure as the (mean) probability of being self-employed over the individuals’ observation period.
Trust in Strangers	2003, 2008	The question on interpersonal trust reads as follows: “If one is dealing with strangers, it is better to be careful before one can trust them.” Response options were given on a four-point scale, allowing the respondents to totally or slightly agree, or totally or slightly disagree with the given statements. Following Glaeser et al. (2000), we define a dichotomous variable by grouping the former and latter two answers.
Unemployment Duration	1992–2010	In every year, respondents are asked to indicate the number of months spent in (registered) unemployment. We calculate our measure of unemployment duration as the average number of months in unemployment per year over the sampling period.
Control Variables		The set of control variables includes the respondents’ age and gender.
Panel C – County-Level Data (See Section 2.4.2)		
County Size	1990	The area of each East German county is reported in Rudolph (1990).

continued

Table 2.B.1 continued

Variable	Years	Source
Demographics	12/1989	Information on age-specific population shares are obtained from infas (*DDR-Kreisstrukturdaten*).
Industry Controls	09/1989	Information on the goods value of production is collected from infas (*DDR-Kreisstrukturdaten*). Data on the industrial composition of the workforce as of September 1989 is reported in Rudolph (1990).
Opposition	1953	We use cartographic statistics published by the former West German Federal Ministry of Intra-German Relations (*Bundesministerium für gesamtdeutsche Fragen*) to construct our measures of regional opposition to the regime. The map was taken from the archives of the Federal Foundation for the Reappraisal of the SED Dictatorship (*Bundesstiftung zur Aufarbeitung der SED-Diktatur*), signature: EA 111 1889.
Political Ideology	1928–1932	We proxy historic political ideology by the mean vote shares for the Communist party (*Kommunistische Partei Deutschlands, KPD*) and the Nazi party (*Nationalsozialistische Deutsche Arbeiterpartei, NSDAP*) in the federal elections in 1928, 1930, 07/1932 and 11/1932 to construct two distinct measures of political ideology. Data on Weimar Republic election results are taken from King et al. (2008).
Population	1925–1933	Population figures for the Weimar Republic were obtained from King et al. (2008) and Falter and Hänisch (1990).
	1980–1989	Data were collected from the Statistical Yearbooks of the German Democratic Republic (*Statistische Jahrbücher der Deutschen Demokratischen Republik*).
Religious Composition	1925	The respective population shares of Protestants and Jews is based on information from the 1925 census of the Weimar Republic (*Volkszählung 1925*). Our data stems from King et al. (2008).
Socialist Indoctrination	1988	We proxy regional socialist indoctrination by the share of political and economic elites that were members of the Socialist Unity Party (SED). We calculate this measure using data from the Central Cadre Database (*Zentraler Kaderdatenspeicher, ZKDS*). This large administrative data set was used for planning purposes and contains information on all political and economic executives of the GDR (except for employees of the Ministry for State Security, the Ministry of National Defence and the Ministry of Internal Affairs). The dataset was obtained from the Federal Archives (*Bundesarchiv*), signature: DC 20 MD/1.
Workforce Composition	1925, 1933	County-level self-employment shares are based on the 1925 and 1933 censuses of the Weimar Republic (*Volks- und Berufszählung 1925 und 1933*). Data for 1925 are obtained from Falter and Hänisch (1990); data for 1933 from King et al. (2008). The share of white-collar workers is based on data from the 1933 census.

Table 2.B.2: Descriptive Statistics

	Mean (1)	SD (2)	P25 (3)	P50 (4)	P75 (5)	Min (6)	Max (7)	N (8)
Panel A – Spying Intensity								
County-Level Spying Density	0.37	0.13	0.26	0.35	0.45	0.12	0.78	148
District-Level Spying Density	0.37	0.07	0.30	0.38	0.42	0.27	0.50	147
District-Level Spying Density (Leave-Out)	0.37	0.07	0.31	0.38	0.43	0.26	0.52	147
Panel B – Dependent Variables								
Trust in Strangers (2003, 2008)	0.12	0.32	0.00	0.00	0.00	0.00	1.00	3,307
Reciprocal Behavior (2005, 2010)	0.77	0.12	0.69	0.76	0.86	0.36	1.00	2,946
Attend Elections (2005, 2009)	0.78	0.27	0.60	0.80	1.00	0.20	1.00	3,041
Extreme Left/Right Voter (2005, 2009)	0.08	0.27	0.00	0.00	0.00	0.00	1.00	3,062
Unemployment Duration 1992–2010	0.11	0.19	0.00	0.00	0.16	0.00	1.00	2,795
Log Mean Labor Income 1992–2010	7.38	0.50	7.12	7.40	7.68	4.92	9.32	2,308
Self-Employment 1992–2010	0.01	0.08	0.00	0.00	0.00	0.00	1.00	2,792
Panel C – Control Variables								
Age (in 1990)	42.44	15.81	30.00	40.00	54.00	16.00	95.00	3,229
Male	0.48	0.50	0.00	0.00	1.00	0.00	1.00	3,229
Stasi On-Site Office	0.03	0.16	0.00	0.00	0.00	0.00	1.00	148
Log Mean Population 1980–1988	11.13	0.55	10.74	11.09	11.42	9.94	13.22	148
Log County Size	5.92	0.80	5.57	6.12	6.52	3.26	7.14	148
City County	0.16	0.37	0.00	0.00	0.00	0.00	1.00	148
Share of Population Aged under 15, 1989	19.64	1.76	18.68	19.57	20.74	15.56	24.74	148
Log Industrial Output 1989	21.11	1.33	20.37	21.30	22.03	16.99	23.73	148
Share Industrial Employment 09/1989	46.75	12.75	37.70	48.05	56.25	16.80	74.50	148
Share of Cooperative Workers 09/1989	12.91	8.90	5.97	11.01	18.16	1.20	35.91	148
Uprising Intensity 1953: Strike	0.22	0.42	0.00	0.00	0.00	0.00	1.00	148
Uprising Intensity 1953: Demonstration	0.17	0.38	0.00	0.00	0.00	0.00	1.00	148
Uprising Intensity 1953: Riot	0.14	0.34	0.00	0.00	0.00	0.00	1.00	148
Uprising Intensity 1953: Prisoner Liberation	0.13	0.34	0.00	0.00	0.00	0.00	1.00	148
Uprising Intensity 1953: Military Intervention	0.76	0.43	1.00	1.00	1.00	0.00	1.00	148
Mean Vote Share Communist Party 1928–1932	14.11	6.08	9.64	13.72	17.08	3.13	36.41	148
Mean Vote Share Nazi Party 1928–1932	26.19	3.99	23.47	26.01	28.53	10.19	37.99	148
Share Protestants 1925	91.35	7.68	90.69	92.76	94.23	16.44	97.10	148
Share Jews 1925	0.17	0.17	0.07	0.12	0.22	0.01	1.32	148
Mean Share of Self-Employed 1925 and 1933	18.46	2.95	16.59	18.14	20.51	11.91	26.90	148
Share of White Collar Workers 1933	7.26	2.33	5.69	6.87	8.24	3.31	15.60	148

Notes: This table presents descriptive statistics on outcome and control variables as well as our main explanatory variable. Measures of the spying intensity as well as most of the control variables vary at the county level, whereas outcomes vary at the individual level, which is reflected in the respective numbers of observations. For detailed information on all variables, see Appendix Table 2.B.1.

2.B.2 Redrawn County Boundaries and Data Harmonization

We combine county-level data from various sources and decades in this study. Since 1925, the first data year in our analysis, county borders have been redrawn multiple times. To account for these territorial changes, we harmonize all county-level data to boundaries as of October 1990. Note that this procedure only applies to county-level data. It is not necessary to harmonize county borders in the individual-level SOEP data as we select respondents based on their reported county of residence in 1990 and track these people over time.

As regards data from the time of the GDR, we have to account for minor territorial reforms only. In ten cases, neighboring counties were merged. In five cases, bigger cities became independent counties (*Stadtkreise*) from the surrounding rural county. We manually account for these administrative changes using detailed maps and other historical sources. When merging two counties, we use the maximum value of the observed opposition variables measuring the strike intensity. In the cases where new counties were constituted, we assign historical values of the emitting county to the created one.

When harmonizing data from the Weimar Republic with the county boundaries as of 1990, greater territorial reforms have to be taken into account. Due to the lack of adequate population weighting factors, the harmonization is based on geospatial area weighting factors as described in Goodchild and Lam (1980). We overlay the corresponding GIS shapefiles from the Weimar Republic with the shapefile from 1990 and calculate area weighting factors that allow for adjusting the historical data to county borders as of 1990. MPIDR and CGG (2011) provide a rich set of historical shapefiles for the German territory. Given that most of our outcomes and control variables refer to people and not space, it needs to be stressed that this procedure is subject to some degree of imprecision. Given the long time span, the numerous territorial reforms, and the lack of population weighting factors, this procedure is, however, the most accurate harmonization strategy we can apply.

Chapter 3

Asymmetric Labor-Supply Responses to Wage-Rate Changes*

3.1 Introduction

Because the labor-supply literature typically focuses on *marginal* wage changes, a common prediction of theoretical models is that labor supply responds symmetrically to wage increases and decreases. In other words, wage increases and wage decreases of equal magnitude have the same effect (though with opposite signs) on labor supply decisions, implying that labor-supply elasticities with respect to wages do not depend on the sign of the wage variation. However, it is not difficult to show that *non*-marginal wage changes, which are the more relevant types of wage changes in the real world and therefore for empirical analysis, can lead to asymmetric responses.[1] Although this result has important implications for the empirical estimation of labor-supply responses, there is little empirical evidence regarding the symmetricity of the effect of wages on labor-supply.

This paper contributes to the literature by estimating the symmetricity of labor-supply responses to non-marginal wage changes. Our precise research question is: do wage increases and decreases of equal magnitude have symmetric effects on labor supply? Answering this research question requires a set-up that introduces (quasi-) randomly assigned wage increases and decreases at the same time for comparable individuals. Finding such types of experiments in 'natural' settings is difficult, if not impossible, and thus may partly explain the sparse literature on the symmetricity of labor supply responses to nominal wages.

We address these empirical challenges using a field experiment on labor supply where we randomly assign wage increases and decreases of equal magnitude to workers. Specifically, we set-up a real labor task and invite workers to work on this task in an actual online-labor-market,

* This chapter is based on a revised version of P. Doerrenberg, D. Duncan, and M. Löffler (2016). "Asymmetric Labor-Supply Responses to Wage-Rate Changes: Evidence from a Field Experiment". *IZA Discussion Paper* 9683.

[1] We show below that several models predict asymmetric responses to non-marginal wage increases and decreases; in particular, a standard labor-supply model and a model of loss aversion would predict asymmetry.

namely Amazon's Mechanical Turk (henceforth mTurk). The labor task is advertised on the mTurk website as any other labor task and workers receive wages that are comparable to other wages on mTurk. In addition, the labor task is designed to be perceived as realistic as possible; it requires workers to transcribe scanned German-language documents. Importantly, the workers in our experiment do not know that they are participating in an academic experiment. We announce a certain wage per transcribed picture in the advertisement of our task on mTurk and workers complete a batch of six transcriptions for the wage announced in the mTurk advertisement.[2] After transcribing the first batch of images, all workers are randomly assigned to one of three groups: (i) the wage remains constant (control group), (ii) the wage increases by 20 %, (iii) the wage decreases by 20 %. After the updated wages have been presented to workers, they can select to either stop working on our labor task or keep working as much as they wish. We identify the symmetricity of the labor-supply response by comparing labor-supply behavior between the three randomly assigned groups.

The results show that wage increases have a positive effect on labor supply while wage decreases reduce labor supply, providing clear support for a positive relationship between labor supply and wages. However, the labor-supply response to wage increases and decreases is asymmetric. This asymmetry is especially strong on the extensive margin, defined as the share of workers who quit our task conditional on seeing the treatment information. The estimated extensive-margin treatment effect for workers who experience a wage decrease is approximately twice that of workers who experienced a wage increase. Estimates of the intensive margin response are also suggestive of an asymmetric response where increases have smaller effects than decreases; differences in intensive-margin responses to wage increases and decreases are large, but imprecisely estimated. Our results further show that the wage changes did not have any effect on the quality of transcriptions, which is above 96 percent in all groups.

We discuss several mechanisms that help to rationalize our results regarding the asymmetry of labor-supply effects of wages. First, our results are consistent with standard labor-supply preferences. For example, there might be a positive number of workers in our task with a reservation wage that is between the initially announced wage and the wage in the decrease group; workers in this part of the distribution of reservation wages would quit the task once they learn about the wage decrease. Concurrently, the labor-supply curve might have a particular shape which induces asymmetric responses to *non*-marginal wage-rate changes. Second, the empirical findings are also consistent with a model of loss aversion where the reference wage is equal to the expected wage of \$ 0.15 and the labor supply curve is kinked at this reference

[2] As a result of this design feature, we induce an exogenously determined expectation regarding the per-unit wage throughout our labor task; workers expect the wage to stay constant at the wage which is advertised on the mTurk website and paid for the first six transcriptions. The field experiment therefore allows us to study the labor-supply responses to unanticipated wage changes.

wage. In such a model, wage decreases are predicted to have a larger labor-supply effect than a wage increase of equal magnitude. Third, previous research by Kube et al. (2013) shows that asymmetric labor-supply responses can be explained by reciprocity (see below). We argue that this is an unlikely explanation for our results; since workers are paid per-unit wages, shirking (as a punishment for wage decreases) or supplying extra hard effort (as a reward to wage increases) is not possible in our context. We also rule out treatment induced skill-composition changes across the three groups due to skill-based exits as a possible explanation for our results. In particular, we do not observe that unskilled workers are more likely to exit in the wage-decrease group than unskilled workers in the other groups.

Our paper contributes to the literature on labor-supply effects of wage changes. Economists have explored the effect of wages on labor supply for several decades (see Keane, 2011, for a survey). Many of these studies use panel-data sets and exploit positive and negative variation in wages to estimate the wage elasticity of labor supply.[3] Because the elasticity estimated by these studies represents roughly an average of wage-increase-induced and wage-decrease-induced elasticities, our results suggest that existing estimates likely overestimate the effect of wage increases while underestimating the effect of wage decreases. Relatedly, our results further raise questions about the comparability of labor-supply elasticities across studies that differ in the sign of the wage changes used for identification. Our findings suggest that it cannot be concluded from the estimated elasticities that workers are more responsive in the one setting relative to another without knowing whether the sign of the wage changes is the same.[4]

Our finding that the largest asymmetry is along the extensive margin is especially important for understanding the labor-supply effects of wages since it is generally accepted that labor-supply elasticities are mainly determined by the extensive margin response (Blundell and MaCurdy, 1999, Meghir and Phillips, 2010, Bargain et al., 2014). Our results highlight one possible reason for the downward rigidity in nominal wages (Kaur, 2018). Among the explanations for this rigidity are the potentially detrimental effects on productivity and labor supply. We find large negative extensive margin responses, which suggest that nominal wage cuts could be damaging for firms – one potential reason for the reluctance of firms to reduce wages.

[3] It is sometimes argued that nominal wage cuts are rare and therefore not relevant. While we acknowledge that nominal wage cuts occur less often than increases (see the literature on nominal wage rigidities, e.g., Kaur, 2018), it has been shown that wage cuts do happen; for example during recessions and bankruptcies, and for the self-employed and salary earners (Kahn, 1997). In addition, many studies on labor-supply elasticities use upward and downward variation in tax rates to instrument for wages (e.g., Eissa and Liebman, 1996, Rothstein, 2010). This generates downward variation in wages even in the absence of nominal wage cuts. Our study is also relevant for decreases in real wages, which occur more frequently than nominal wage cuts. Our results suggest that inflation-induced decreases of real wages have larger labor supply effects than previously thought.

[4] This is especially important for meta-analysis studies on labor supply (e.g., Evers et al., 2008). Our findings imply that in such meta-analyses one should carefully distinguish between labor supply estimates based on wage increases and those based on wage decreases.

We further add to the experimental literature on the effect of wages on effort and labor supply. These studies provide credible randomized evidence in the absence of (discrete) work-time constraints, something which is difficult to obtain using observational data. Papers based on laboratory experiments provide robust evidence that labor effort and wages are characterized by a positive relationship (see the survey by Charness and Kuhn, 2011), which is consistent with our findings. However, laboratory experiments are subject to the usual concern that they cannot easily be generalized to real-world situations. Field experiments with higher external validity find mixed effects regarding the relationship between wages and effort. While some field experiments find a positive effect of wages on effort/labor supply (Fehr and Goette, 2007, DellaVigna and Pope, 2018), other studies find either no relationship (Hennig-Schmidt et al., 2010), short-run temporary effects which do not make a difference for final work outcomes (Gneezy and List, 2006), or (positive) effects for only certain types of workers (Cohn et al., 2015). Our results add to the (ongoing) discussion on the wage-effort relation in field experiments, and provide evidence of a positive relationship between wages and labor effort in online labor markets.

To the best of our knowledge, no (lab or field) experimental study explores the potentially differential effects of wage increases and decreases on labor supply.[5] An exception is Kube et al. (2013), which is the study most closely related to ours. They conduct a field experiment with students working in a library for a given period of time (six hours). They generate an exogenous reference wage by announcing a projected hourly wage to all workers when the job is advertised. Immediately before the task starts, they announce a higher wage to workers in one treatment group and a lower wage to workers in another group. Workers in the control condition receive the initially announced wage. The study finds that the wage cut decreases work effort (i.e., output generated during the given period of time) whereas the wage increase does not have any effect relative to the control group. In line with our findings on transcription accuracy, their study also does not find any effects on quality of work.

While these results are broadly consistent with our findings, our paper differs from theirs in the design of the labor market institution, which has important implications for the interpretation and application of our findings. The institutions differ in that we pay workers for each transcribed picture instead of for a predetermined number of hours, and we allow workers to quit the labor task whenever the choose to do so. Furthermore, our analysis is based on a much larger sample of workers from a real-world labor market. Therefore, our design is representative of labor markets where workers receive piece-rate payment and have tremendous labor supply flexibility, whereas Kube et al. (2013) focus on labor markets where workers are required to

[5] Wage cuts are generally understudied in this literature; note that none of the experimental papers referenced above examines wage cuts. Also see footnote 3 regarding the prevalence of wage cuts.

work a predetermined number of hours for a fixed hourly wage rate. One advantage of our design is that workers are able to respond on two additional margins that are not included in Kube et al. (2013); the extensive margin and the intensive-*time* margin.[6] As a result, we are able to study asymmetric responses to wage changes on both margins. Additionally, because our workers receive a piece rate, subjects who reduce output earn a lower pay-off and have less scope to punish their employer through shirking. This implies that we do not study 'work morale' and it reduces the likelihood that our findings are driven by reciprocity as in Kube et al. (2013). Therefore, we are able to show that labor supply asymmetry exists even in the absence of a motive to reciprocate. The institutional frame-work of our study – large sample of workers in their natural labor-market environment – also implies that our findings can be generalized to similarly situated labor markets; large crowd-sourcing labor markets characterized by low wage and high flexibility.[7]

To the extent that our findings can (partly) be explained by a model of loss aversion, our paper further makes a contribution to the behavioral-economics literature on loss aversion following Kahneman and Tversky (1979).[8] There is a large empirical literature showing that individuals indeed have preferences consistent with loss aversion and that individual expectations determine the reference point (e.g., Dunn, 1996, Post et al., 2008, Abeler et al., 2011, Card and Dahl, 2011, Ericson and Fuster, 2011, Pope and Schweitzer, 2011), but there is scarce evidence regarding the role of loss aversion for labor-supply responses. We add to this literature in that we provide evidence that individuals may have preferences that are consistent with loss aversion in the context of labor supply. This finding is consistent with Ahrens et al. (2014) who derive the theoretical prediction that labor supply responds asymmetrically to wage rate changes in a framework with reference-dependent utility functions.[9]

[6] The participants in Kube et al. (2013) work for a pre-specified time period, thus precluding the possibility to study a time response. In our experiment, the participants can choose for how long they work, allowing us to study the intensive-time margin.

[7] The findings of two additional papers are relevant in the context of asymmetries in labor markets. Falk et al. (2006) find in a laboratory experiment that reservation wages respond asymmetrically to the introduction and removal of minimum wages. The results show that the introduction of a minimum wage has larger effects than the removal. Chemin and Kurmann (2014) study how reciprocal behavior of 12 fieldworkers responds to wage increases and decreases. Consistent with Kube et al. (2013), they find that wage increases had no effect while wage decreases had a negative effect on effort, and attribute this effect to reciprocity.

[8] This literature pursues the idea that individuals evaluate outcomes relative to reference points. These types of preferences are commonly termed reference-dependent preferences and have been formalized by Kőszegi and Rabin (2006, 2007, 2009). Loss aversion describes the notion that individuals weight negative deviations (losses) from the reference point more than gains of equal magnitude.

[9] Our paper also relates to several studies showing that individual labor supply decisions are affected by target incomes. In a survey of the literature, Goette et al. (2004) show how empirical results on labor-supply behavior are consistent with reference-dependent preferences where workers provide high effort if they are below a target income, whereas they provide less effort if they have surpassed a target (also see Camerer et al., 1997, Crawford and Meng, 2011 and Fehr and Goette, 2007). Empirical evidence also suggests that loss aversion affects job searches (DellaVigna et al., 2018). While these studies demonstrate that workers have target incomes and

The paper is organized as follows. Section 3.2 describes the real labor task and its implementation in Amazon's Mechanical Turk. Section 3.3 describes the data and our empirical approach and we present the results in Section 3.4. We discuss the potential economic mechanisms behind our findings, as well as their implications and generalizability, in Section 3.5. Section 3.6 concludes.

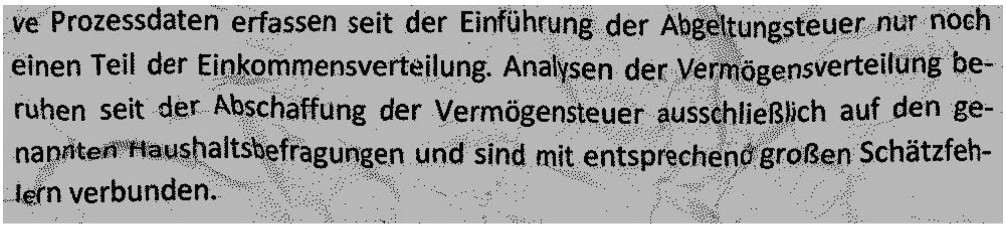
3.2 The Experiment

This section describes the field experiment used to estimate the impact of wage rate changes on labor supply. We begin by describing the labor task, the treatment design and the implementation in Amazon's Mechanical Turk.

3.2.1 Design

Labor Task. We selected an online labor task that requires subjects to transcribe German text shown in a series of images. The German texts are taken from a recent publication, but each page of the document is ruffled so that the scanned versions appear much older than they really are. The advantage of changing the appearance of the images is that subjects are more likely to believe that the texts were scanned from old books for which a digital copy is not available. The task then, is to digitize these "old" German books.[10] Each image has approximately five lines and 43 words (344 characters). Figure 3.1 shows an example. Subjects are randomly shown one of 128 images at a time and are instructed to hit "save picture" when they are done transcribing the text in the image. A new image is shown after the subject hits "save picture".

Figure 3.1: Image of Text to be Transcribed


ve Prozessdaten erfassen seit der Einführung der Abgeltungsteuer nur noch einen Teil der Einkommensverteilung. Analysen der Vermögensverteilung beruhen seit der Abschaffung der Vermögensteuer ausschließlich auf den genannten Haushaltsbefragungen und sind mit entsprechend großen Schätzfehlern verbunden.

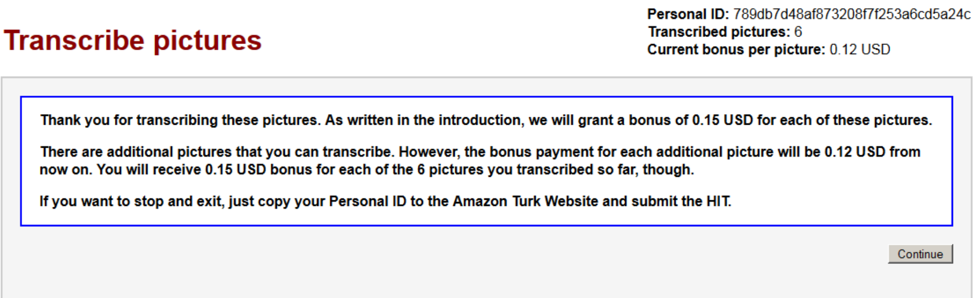
Notes: This figure depicts a screenshot of an image of text that was to be transcribed by the subjects. Subjects were randomly shown one of 128 images. All images are comparable to the image depicted in the figure. All images are in German and taken from a recent policy-report publication.

loss-aversion preferences in the context of labor supply, they do not allow conclusions about the asymmetric effects of wages.

[10] Horton et al. (2011) use a similar task and motivate it with the following advantages: transcribing text (i) is tedious, (ii) requires effort and attention, and (iii) has a clearly defined quality measure.

Treatment Groups. We use a between-subjects design in order to identify the effect of wage changes on labor supply. Subjects are randomly assigned to one of three groups: one control group and two treatment groups. Subjects in all three groups work on the labor task described above and are paid a piece rate for each image that is transcribed. The piece rate[11] is set at \$ 0.15 for each of the first six transcribed images in all three groups. Subjects receive a notification thanking them for transcribing the images after the first six images have been transcribed. They are then told that they can transcribe additional images and that the piece rate for the additional images is either \$ 0.18, \$ 0.15 or \$ 0.12, for the wage-increase, control, and wage-decrease groups, respectively (see Figure 3.2 for an explanatory treatment notification). Notice that the wage rate remains fixed at \$ 0.15 for the control group, and that the wage rate change is the same for both treatment groups; in each case the rate changes by \$ 0.03 or 20 %. We did not provide workers with a reason for their wage changes in order to keep a neutral framing (Kube et al., 2013). In addition, the reasons for the wage changes would have had to be different for wage increases and decreases, which would have complicated the comparability between the treatment groups.

Figure 3.2: Treatment Variation



Notes: This figure depicts a screenshot of the treatment notification in the "wage decrease" group. The treatment notifications for the "control" and "wage increase" groups were identical except for the information regarding the piece-rate wage for the subsequent images. The treatment notification popped up after a subject transcribed six images.

Wage Expectations. The experiment is designed to establish an exogenous and salient expectation regarding the per-unit wage in the our mTurk task. Potential workers are told that the wage per transcribed picture is \$ 0.15 in the job announcement. Additionally, workers who start working on our task face the announced wage of \$ 0.15 for the first six transcribed pictures, after which the wage rate either increases or decreases. We argue that this design generates the expectation that the per-unit wage will remain constant at \$ 0.15 throughout the entire

[11] The piece rate is called *bonus* in the experiment. This is the usual wording if one is to implement per-piece payment within the same task in the mTurk labor market.

task. Our experimental design therefore allows us to study how unexpected wage increases and wage decreases affect labor supply. If we had initially told subjects that the wage would either increase or decrease, they could have adjusted their expectations and the labor supply response to varying wages would not have been comparable to real-situations where workers experience unanticipated wage changes. This design feature is also in accordance with Kube et al. (2013) who, following Bewley (2005), argue that deviations from an exogenous expectation capture the key aspects of wage changes (for example, disappointment and the break of trust relation in the case of wage cuts).

One potential drawback of our experimental design is that it may raise concerns of deception since the job description does not notify subjects of the possibility that the wage may increase or decrease after a certain number of transcribed pictures. This was a deliberate choice in an effort to establish a clear and salient wage expectation.[12] We avoid deception by including the following pieces of information in the treatment notification (see Figure 3.2). First, we thank the workers for completing the transcription task and remind them that, as promised in the introduction of the task, they will be paid $ 0.15 for each of the pictures they transcribed so far. Next, we inform them that they have the option to transcribe additional images and that the piece rate for these additional transcriptions is different from that for the first batch of transcriptions. Finally, we make it clear that they can stop and exit the task at this point if they wish and instruct them on what to do next to ensure we are able to process their payment.[13] We argue that these design features make it clear to workers that they first transcribe pictures based on the piece rate announced in the introduction to the task, and that they can transcribe additional pictures at a new rate. The design of the task gives the impression to workers that the task consists of two parts and ensures that we did not deceive the workers regarding the wage in the second task.

3.2.2 Implementation

Labor Market and Recruitment. The experiment is implemented in the field using workers on Amazon's Mechanical Turk. mTurk is an online labor market where job offers are posted and workers choose jobs for payment. It has numerous benefits for running experiments, including

[12] Informing subjects about the possibility of a wage change would have generated uncertainty about the eventual wage and the wage expectation would not have been as clear.

[13] The notification reads: "Thank you for transcribing these pictures. As written in the introduction, we will grant a bonus of $ 0.15 for each of these pictures. There are additional pictures that you can transcribe. However, the bonus payment for each additional picture will be $ 0.12/$ 0.18 from now on. You will receive $ 0.15 bonus for each of the six pictures you transcribed so far, though. If you want to stop and exit, just copy your Personal ID to the Amazon Turk Website and submit the HIT." Instead of the wage change, we include the following message for the control group: "There are additional pictures that you can transcribe. Just as before, the bonus for each additional picture will be $ 0.15."

access to a large stable subject pool, diverse subject background, and low cost.[14] Furthermore, the behavior of online workers has been shown to be comparable to those of subjects in laboratory studies (Horton et al., 2011). Additionally, experimenter effects are avoided because subjects do not know that they participate in an experiment (Paolacci et al., 2010, Horton et al., 2011, Buhrmester et al., 2011, Mason and Suri, 2011). Importantly for us, we are able to identify the effect of wages changes in a naturally occurring labor market. In general, experiments on Amazon's Mechanical Turk therefore combine internal and external validity since it is a real labor market with actual workers where randomized trials can be conducted (Horton et al., 2011).[15]

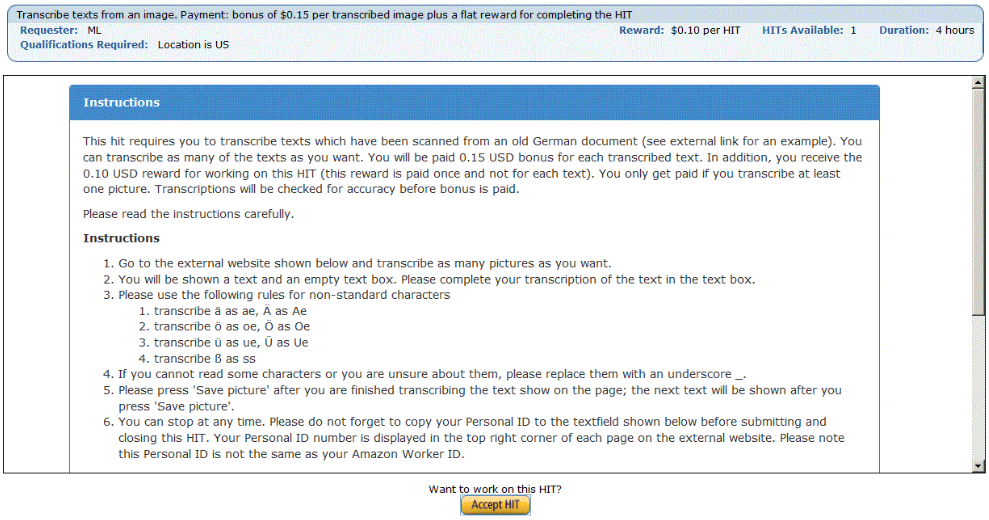
Although we recruit workers through mTurk, they complete the labor task on an external website that we created for the purposes of the experiment. We first create a human intelligence task (HIT) that is advertised on mTurk. The HIT includes a description of the labor task and compensation. It also includes instructions for how to complete the task; see Figure 3.3. Particularly, subjects are told to accept the HIT and click on the weblink if they are interested in completing the task. Subjects who click on the link are taken to our external website where they are randomly assigned to one of three groups and shown the instructions in Figure 3.4. Subjects are instructed to click continue if they wish to work on the task, and those who do are shown images of scanned German text that they must transcribe for payment. Each page of our website shows the subjects their personal ID, number of pictures transcribed so far, and the current piece rate. We implement treatment after six images have been transcribed and limit the total number of images that each subject can transcribe to 50. However, subjects are not aware of either of these limits until they reach them. In other words, subjects do not know that the HIT has six images, that they will have the opportunity to continue working after the first six images, that the piece rate might be different if they continue working, or that they can only transcribe up to 50 images if they chose to continue working. Subjects in wage-decrease group who complete six transcriptions are shown the treatment information illustrated in Figure 3.2. A similar text is shown to subjects in the wage-increase group and the control group; the only difference is the piece rate for the additional images.

Transcribing text from an image can be a tedious task. However, given that the text in the images is short, the task could be perceived as mostly costless for German speakers. In order to reduce this possibility and ensure that the labor costs are non-zero, we restrict the subject pool to workers with a US IP address. The idea here is that the labor cost of transcribing German text is much higher for non-Germans than for Germans. Of course, our restriction does not preclude

[14] According to Amazon, there are over 500,000 workers from 190 countries in the mTurk labor market: `https://requester.mturk.com/tour`.

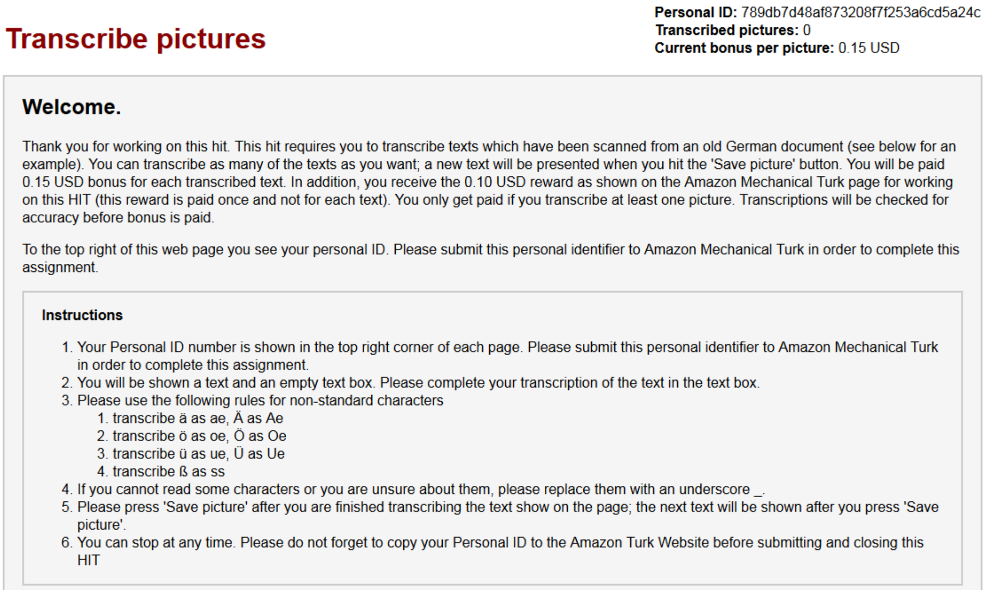
[15] Kuziemko et al. (2015) and DellaVigna and Pope (2018) are recent examples of economics papers using Amazon's Mechanical Turk.

Figure 3.3: Human Intelligence Task Shown on mTurk



Notes: This figure depicts a screenshot from Amazon's Mechanical Turk website. It shows how the labor task used for the field experiment was advertised on mTurk. Subjects are taken to our external website once they click the "Accept Hit" button.

Figure 3.4: Instructions Shown on Our Website



Notes: This figure depicts a screenshot of the external website that we set up for the purpose of the field experiment. Subjects were taken to this website once they decided on Amazon's Mechanical Turk website that they would like to work on the task. The depicted screenshots provides subjects all information relevant for the task.

the possibility that German speakers participated in the task. However, any Germans who participated in our experiment are randomly distributed across our treatments and therefore have no effect on our outcome of interest.

The experiment is programmed on mTurk to expire after 750 workers accept the HIT or 10 days have passed, which ever comes first. Our initial run of the experiment, which started on June 15, 2015, expired after 10 days with only 418 workers. Therefore, we initiated a second run on July 20, 2015, and this run expired after hitting the 750 worker threshold six days later. In total, 1,168 workers participated in the two runs. Note that the HIT is designed such that workers cannot work on the task more than once. We also excluded workers who participated in the first run from participating in the second run. Moreover, it is highly unlikely that individuals have multiple worker accounts to avoid these constraints: First, when registering for mTurk, Amazon requires workers to confirm in the Participation Agreement that they "may not use multiple Amazon Accounts to register with Mechanical Turk". Second, the Participation Agreement further requires workers to provide "true and accurate" information on a worker's name, email address, phone number and physical address.[16] Third, workers are required to provide a tax identification number (Social Security Number or Individual Tax Identification Number) after their mTurk lifetime earnings have exceeded a set threshold. Workers who fail to provide this number are not allowed to accept additional HITs on mTurk.

Payment. The experiment ends for each subject when she decides to stop or when she transcribes 50 pictures, whichever comes first. In either case, each subject is instructed to copy her personal ID number, which is shown in the top right corner of each page, and paste it in the entry box on the mTurk website. This process is necessary for us to match subjects to their mTurk worker ID and thus process their payments. Subjects receive a participation reward of $ 0.10, which is paid as long as a subject accepts the HIT and completes at least one transcription. Additionally, subjects are paid a piece rate of $ 0.15 for each of the first six transcribed pictures, and depending on treatment group, $ 0.12, $ 0.15 or $ 0.18 for each transcribed image above the first six transcriptions. Given the payment restrictions imposed by the mTurk platform, we frame the piece rate as a bonus in all communications to the subjects. For example, subjects in the control group are told they will be paid $ 0.10 for participating in our HIT and a bonus of $ 0.15 for each transcribed picture.

We chose this payment structure based on a small test of the real effort task that we implemented with English-speaking students in a class at a major public university in the US before we started the field experiment.[17] This test revealed that it takes about 4 minutes on average to

[16] The Participation Agreement is online at: `https://www.mturk.com/mturk/conditionsofuse`.

[17] This test did not include any wage variations. The sole purpose was to test the functionality of the website and

transcribe foreign-language text paragraphs that have the same size as the paragraphs in our experiment. This suggests that approximately 15 pictures can be transcribed per hour, resulting in an hourly wage of about \$ 2.35 ($= 0.1 + 6 \times 0.15 + 9 \times 0.15$) in the control group, \$ 2.62 ($= 0.1 + 6 \times 0.15 + 9 \times 0.18$) in the increase group, and \$ 2.08 ($= 0.1 + 6 \times 0.15 + 9 \times 0.12$) in the decrease group. In light of a median reservation wage of between \$ 1.12 and \$ 1.38 per hour for mTurkers, according to Horton and Chilton (2010) and Horton et al. (2011), this payment structure seemed adequate from an ex-ante perspective. From an ex-post perspective (see results), it turns out that the average time needed per picture in our small test was an appropriate, if not even too conservative, predictor of the transcription speed in our actual experiment. We observe that participants in our sample (across all groups) needed about 3.7 minutes per picture (i.e., about 16.22 pictures per hour), which results in hourly wages of \$ 2.53, \$ 2.84, \$ 2.23 in the three groups, respectively. That is, both the per-hour wages that we expected before we ran the experiment and the per-hour wages that we observe for the workers in our experimental sample are considerably higher than the hourly median reservation wage reported in Horton and Chilton (2010) and Horton et al. (2011).

3.3 Data and Empirical Approach

This section describes our outcome variables, details on the sample, and the empirical strategy used to identify the symmetry of wage effects.

3.3.1 Outcome Variables

We construct several outcome variables that measure different aspects of labor supply in order to identify the effect of wage changes on labor supply. These include the quit rate, number of transcribed pictures, time spent transcribing, transcription rate, and accuracy. Each of these variables is described in greater detail below.

Transcriptions and Hours. Because workers are paid for each transcribed image, we expect that they will respond to the wage changes by changing the number of images they transcribe. Therefore, one variable of interest is the total number of transcribed images per worker. We further explore the total time spent working on the task and the time per transcribed text (transcription rate). Because we do not have an exact measure of the time workers actually spent working on a picture, we proxy the transcription rate by counting the time between the submission of two transcriptions. We acknowledge that this likely overstates the transcription

to infer the average time it takes to transcribe one of the images.

time for any given image. However, the difference in transcription rate between groups should still be instructive of the impact of wage changes.

Extensive Margin. Recall that workers are notified of treatment after transcribing six images. The notification makes it clear that the worker has completed the HIT, but that there are additional (optional) images to transcribe. Workers are also informed that they can quit the task at this point or continue transcribing the additional images at the newly announced wage rate. Given these features of the treatment notification, we interpret the decision to stop working at this point as an extensive margin decision. Therefore, one of our key outcome variables is the share of workers who quit the task immediately after receiving the notification. Because the treatment notification has a modest nudge to quit, we expect that the share of quitters will be reasonably high in the control group despite the fact that the wage remains constant. The important question for us is: does the wage increase/decrease have any effect beyond this modest nudge.

An important feature of online-labor markets such as mTurk is that they facilitate almost instantaneous switching of labor tasks. In other words, a worker can quit one job this second and start a new job the next second. This is not unlike what one would observe in traditional labor markets where a worker secures a new job before quitting her existing job. Unfortunately, we do not observe what subjects do when they quit our task. Therefore, the extensive margin response in our study simply means that the worker quits our task. We cannot say whether or not they quit working online or switch to a more profitable task.

Accuracy. Recall that the transcriptions are based on text for which we have the original digital copy. This makes it possible for us to measure accuracy by comparing the transcribed text for each worker to the actual text.

3.3.2 Sample

Our HIT was accepted by 1,168 mTurk workers. We restrict the sample to those workers who completed at least one picture, and therefore received the participation fee; this leaves us with 1,158 workers. We observe in the data that a few workers worked on the task for an unreasonable number of time, e.g., several days. To avoid this source of noise, we drop the top 0.05 % of workers in the distribution of minutes worked; these are six workers who worked for more than 385 minutes on the task. Table 3.1 presents summary statistics for our sample of workers ($N = 1,152$) with regard to our main variables: number of transcribed pictures, accuracy of transcription, and total time worked. We observe that, on average, workers

Table 3.1: Summary Statistics

Variable	N	Mean	SD	P10	P50	P90
Pictures Transcribed	1,152	12.81	13.23	2.00	7.00	33.00
Total Time	1,152	39.79	50.25	3.17	19.64	104.35
Accuracy	1,151	0.97	0.02	0.96	0.97	0.98

Notes: This table provides summary statistics for outcome variables. The sample is all subjects who started working on the task (i.e., including those who did not necessarily get to see the treatment notification after six transcribed pictures). *Pictures Transcribed* is the average number of images that subjects transcribed. *Total Time* is the average time (in minutes) that subjects totally spent on working on the labor task. *Accuracy* the average share of characters that is transcribed correctly. *N* is the number of observations. *SD* is the standard deviation. *Px* indicates the x-th percentile.

transcribed 12.8 pictures[18] over an average time span of 39.79 minutes. The transcription quality was very high with an average accuracy of 96.97 %. This is reassuring as it suggests that workers take the task seriously and provided high-quality transcriptions. Note that we intended to avoid giving the impression that subjects are participating in an experiment, and therefore did not survey any demographic characteristics.

Because the treatment variation in wages only appears after the first batch of six transcriptions, only a share of the total 1,152 participants are exposed to the treatment condition. Table 3.2 shows that 62.5 % (720) of the 1,152 workers completed at least six pictures and therefore saw the treatment notification. This share ranges from 59 % in the wage-increase group to 65 % in the wage-decrease group. The number of observations in each treatment group is summarized in Table 3.2. In total, we have 248, 215, and 257 workers who saw the treatment notification in the control, increase and decrease groups, respectively. Because workers did not know they were in an experiment or that the wage rate would change, self-selection into the treatments was impossible. We therefore argue that the groups are balanced with respect to the characteristics that predict the probability of quitting before seeing the treatment, and thus we mostly restrict the empirical analysis that follows to the sample of 720 participants who saw the treatment (see Section 3.4.1 for data-based evidence that there is no difference between groups before treatment notification).

A common feature of mTurk is that workers discuss HITs on forums. This can raise issues for experimenters as those workers who have completed the experiment will unknowingly share the details of treatments with other workers who have yet to complete the experiment. We followed the forums on mTurk in order to determine if our HIT was being discussed and discovered that our HIT did in fact show up on one of the forums.[19] The first mention of our HIT occurred on July 24 during the second run of the experiment. We noticed the mention on

[18] Appendix Figure 3.A.1 provides the distribution of completed pictures for all workers in the sample.

[19] See `https://www.reddit.com/r/HITsWorthTurkingFor/comments/3eg39l/us_transcribe_texts_from_an_image_payment_bonus/`.

Table 3.2: Number of Observations

Group	Seen Treatment No	Seen Treatment Yes	Total
Control Group	143	248	391
Wage Increase	149	215	364
Wage Decrease	140	257	397
Total	432	720	1,152

Notes: Number of observations by treatment group who (i) started working on the task but did not see the treatment notification, i.e., they transcribed five images or less (column *No*) and (ii) who transcribed at least six pictures and therefore saw the treatment notification (column *Yes*).

the 26th when the HIT had already expired. The discussion on the forum was favorable towards our HIT, but workers discussed the fact that the wage rate changed as well as the magnitude of the changes. They also discussed potential reasons for rate changes, and mostly speculated that the wage variation must be due the quality of work. Nobody speculated that this task is an experiment; people therefore still did not know they were part of an experiment.

The forum post led to a significant spike in acceptance of our HIT; approximately 58 % of the workers accepted the HIT after the forum discussion began. Because some of these subjects knew of a potential wage variation before accepting the HIT, self-selection might be a problem. For example, it is possible that only workers who are willing to work for our lowest wage rate accepted our HIT. If this is the only source of selection, then our analysis produces a lower bound estimate in both groups. A more troubling source of selection is a case where workers sign up with the hope of receiving a wage increase. These subjects would effectively have the expectation that the wage will be $ 0.18, and would be more likely to quit the task if assigned to the wage decrease group. This source of selection would lead to a downward bias in the wage-decrease group and upward bias in the wage-increase group. Because of this potential problem, we present estimates with and without the post-forum sample. There is no evidence that the forum had an effect on the results (see Appendix Figures 3.A.7–3.A.11).

3.3.3 Empirical Strategy

Random assignment to treatment groups ensures that our empirical approach is straight forward. The empirical analysis proceeds as follows. First, for each experimental group, we plot the share of subjects in each 'period',[20] relative to the total number of subjects who initially started

[20] We use the term 'period' to indicate the number of the picture which is to be transcribed. For example, the treatment notification occurred after 6 periods; i.e., after subjects transcribed 6 pictures.

the working task. We use all available subjects (i.e., not only those who saw the treatment notification) for this exercise. This descriptive analysis sheds light on differential drop out rates across the experimental groups before and after treatment.

Second, for each outcome variable, we compare the means of the respective outcome across treatment groups and use non-parametric Wilcoxon rank-sum tests for differences in distributions between the groups (Wilcoxon, 1945, Mann and Whitney, 1947). In addition, we run simple OLS regressions of the outcome variables on the treatment dummies.

$$Y_i = \alpha + \beta_{Increase}\mathbb{1}(i \in Increase) + \beta_{Decrease}\mathbb{1}(i \in Decrease) + \epsilon_i \tag{3.1}$$

where Y_i is an outcome of interest for subject i, e.g., the number of transcribed pictures. Indicator functions $\mathbb{1}(i \in Increase)$ and $\mathbb{1}(i \in Decrease)$ evaluate to one if worker i is part of the increase group or the decrease group, respectively, and zero otherwise. α is a constant, ϵ_i denotes the unexplained error term. These empirical analyses allow us to identify the effect of wage increases and decreases on our outcome variables. These parametric and non-parametric analyses are restricted to the subjects who saw the treatment notification (this is sufficient because there is no selection prior to the treatment notification – as discussed before and shown below in Section 3.4.1).

To test for symmetry of these responses, we use the $\widehat{\beta}$ coefficients of the above OLS regressions for each outcome and t-tests to test the null that the sum of the estimated coefficients for the wage-increase group and wage-decrease group (both relative to the control group) is zero:

$$H_0 : \beta_{Increase} = -\beta_{Decrease}. \tag{3.2}$$

We then use the estimated regression-based treatment effects to calculate implied elasticities separately for each treatment group. Using the control group as a counterfactual, we derive the elasticity of an outcome variable Y with respect to wages for each treatment group g (either wage increase or decrease) as follows:

$$\epsilon_g = \frac{(Y_g - Y_c)/Y_c}{(w_g - w_c)/w_c} \qquad \widehat{\epsilon}_g = \frac{\widehat{\beta}_g/\widehat{\alpha}}{(w_g - 0.15)/0.15}, \tag{3.3}$$

where subscript c indicates the control group, w is the wage per transcribed picture, $(w_g - w_c)$ is the change in wages in group g (either +3 or -3), and $(Y_g - Y_c)$ is the difference between the outcome variable in group g and the control group. Specifically, $(Y_g - Y_c)$ is the difference in means between the relevant treated group and the control group or, equivalently, the regression coefficient β_g of the respective treatment dummy.

Finally, to shed more light on the dynamics of the effects, we regress the probability of working in a given period on a full interaction of an increase-group dummy and period dummies as well as a full interaction of a decrease-group dummy and period dummies (standard errors clustered on individual level). Formally, this regression reads:

$$D_{i,t} = \sum_{t=1}^{50} \gamma^t_{Increase} \mathbb{1}(i \in Increase) + \sum_{t=1}^{50} \gamma^t_{Decrease} \mathbb{1}(i \in Decrease) + \nu_t + \epsilon_{i,t} \quad (3.4)$$

where $D_{i,t}$ equals one if worker i transcribed picture t and zero otherwise. We normalize coefficients of the last pre-treatment period – $\widehat{\gamma}^6_g$ – for both the increase and the decrease group to zero. The coefficients from this regression, which we present in an event-study type figure, provide insight about the differential dynamics of our treatment effects between the wage-increase and wage-decrease groups (always relative to the control group). Interested in the asymmetry of treatment effects between groups, we then use these regression results to sum up the estimated coefficient for the wage-increase group $\widehat{\gamma}^t_{Increase}$ and the corresponding estimate for the wage-decrease group $\widehat{\gamma}^t_{Decrease}$ for each period t. These results are also shown in an event-study type figure and provide insights about the dynamics of asymmetric effects.

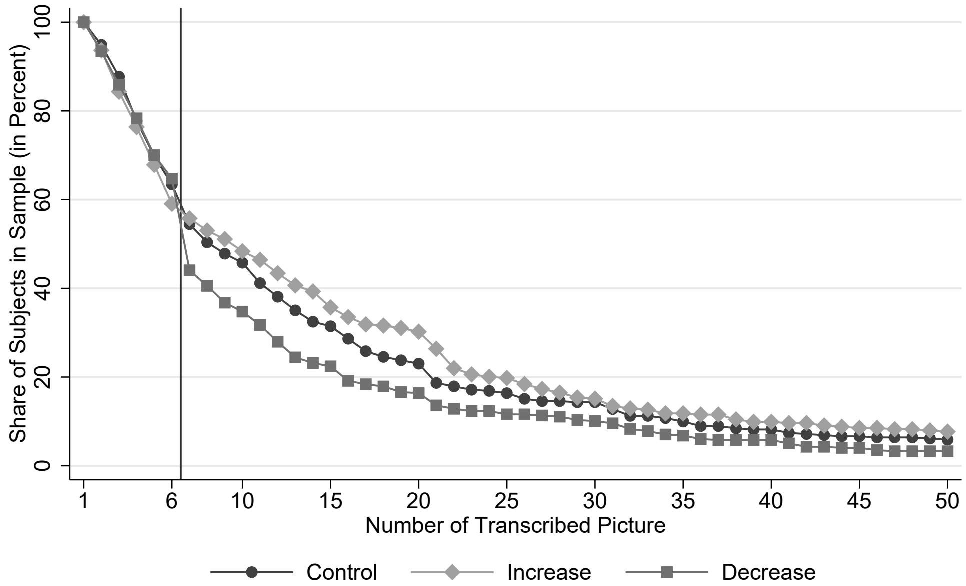
3.4 Results

In this section we present the empirical results of our experiment. We start by analyzing workers' drop-out rates by looking at the full sample with all subjects to check that there are no pre-treatment trends in outcomes in Section 3.4.1. This descriptive exercise also allows to check for first indications of asymmetric responses across the three groups. In a second step, we study the treatment effects at the extensive margin restricting our sample to workers who transcribed at least six pictures and saw the treatment (Section 3.4.2). Third, we analyze intensive margin responses regarding the time spent working and the transcription rate in Section 3.4.3, and on the number of transcribed pictures and the quality of transcriptions (see Section 3.4.4).

3.4.1 Descriptive Evidence

We begin the analysis by calculating the drop-out rate in each 'period' by control and treatment groups. Using all workers who completed at least one picture, Figure 3.5 shows that around 7-8 percent of those subjects who started working in our experiment quit the task in each pre-treatment period. Importantly, Figure 3.5 shows that pre-treatment drop-out rates are equal across groups, suggesting that there is no differential selection across groups *before* treatment. After seeing the treatment, the exit-rate in the increase group decreases to three percent, while the exit rate for the decrease group rises to 21 percent. These differential exit-rates are a first indicator of asymmetric effects between wage increases and decreases.

Figure 3.5: Share of Workers over Periods by Treatment Group



Notes: This figure depicts the share of subjects in each period, relative to the total number of subjects who initially started the task, by experimental group. For example, the value in period ten indicates the share of subjects who complete ten pictures in a given experimental group as a fraction of all subjects who started working on the labor task in this experimental group. Treatment occurred after period six. The sample includes all participants who started working on the task. Number of observations is 1,152.

3.4.2 Extensive Margin

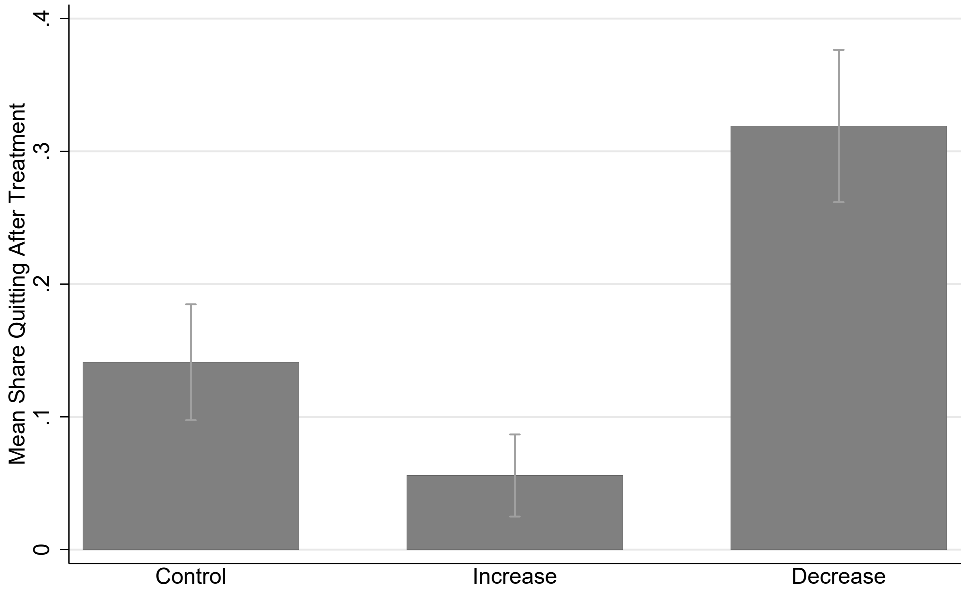
We now turn to the statistical analysis of our results. Because we find no evidence of self selection into treatment – as illustrated by non-differential pre-treatment trends – all following analyses are based solely on the sample of workers who saw the treatment notification.[21]

Figure 3.6 displays the treatment effects on the extensive margin, i.e., the share of workers who quit immediately after having seen the treatment. We observe that 14.1 % of the workers in the control group quit the labor task after receiving the treatment notification. Relative to the control group, the share of quitters is 8.5 percentage points lower in the wage-increase group and 17.8 percentage points higher in the wage-decrease group. These group-wise differences between means are all statistically significant at the 1 % level according to non-parametric

[21] Note that this implies that all subjects in this sample have transcribed at least six pictures and we are left with around 60 % ($N = 720$) of the original sample at this point.

ranksum tests, and suggest that wage increases induce workers to keep working while wage decreases increase the likelihood of quitting. These results are also demonstrated in OLS regressions (based on equation (3.1)) of the extensive-margin indicator variable on the treatment dummies; see Model I of Table 3.3.

Figure 3.6: Extensive Margin by Treatment Group



Notes: This figure depicts the share of subjects in each group who quit the labor task immediately after seeing the treatment notification (i.e., share of subjects who transcribed six pictures but not a seventh one), along with 95 % confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the "wage increase" group and 257 subjects in the "wage decrease" group. All 720 subjects in the sample have transcribed at least six images.

An important observation is that the extensive margin response is asymmetric; the treatment effect for the wage-increase group is economically and statistically different from that for the wage-decrease group (p-value: 0.094, calculated based on equation (3.2)). The asymmetry is also evident in the implied elasticities (as calculated by equation (3.3)), which is 3.0 in the increase group and 6.3 in the decrease group.

We further investigate the dynamics of the treatment effects in Figure 3.7, which graphically presents the coefficients from regression equation (3.4). The left panel shows that our treatment affects subjects throughout the entire experiment until the maximum number of transcribed pictures is reached. In each period after the treatment notification, individuals in the increase group are more likely to still participate in our experiment relative to the control group. Similarly, workers in the decrease group are less likely to continue working compared to the control

Table 3.3: Regression Estimates and Implied Elasticities

	Extensive Margin (1)	Pictures (2)	Total Time (3)	Mean Time (4)	Accuracy (5)
Reference Group: Control					
Wage Increase	-0.085***	3.309**	6.078	-0.344	-0.000
	(0.027)	(1.298)	(4.985)	(0.264)	(0.001)
Wage Decrease	0.178***	-3.795***	-10.556**	0.175	0.000
	(0.037)	(1.166)	(4.832)	(0.358)	(0.001)
Constant	0.141***	19.040***	60.916***	3.778***	0.971***
	(0.022)	(0.870)	(3.399)	(0.210)	(0.001)
N	720	720	720	720	719
R^2	0.082	0.044	0.016	0.003	0.001
$p\ (\beta_{Inc} = -\beta_{Dec})$	0.094	0.820	0.596	0.752	0.906
Elast. Increase	-3.02	0.87	0.50	-0.45	0
Elast. Decrease	-6.30	0.99	0.87	-0.23	0

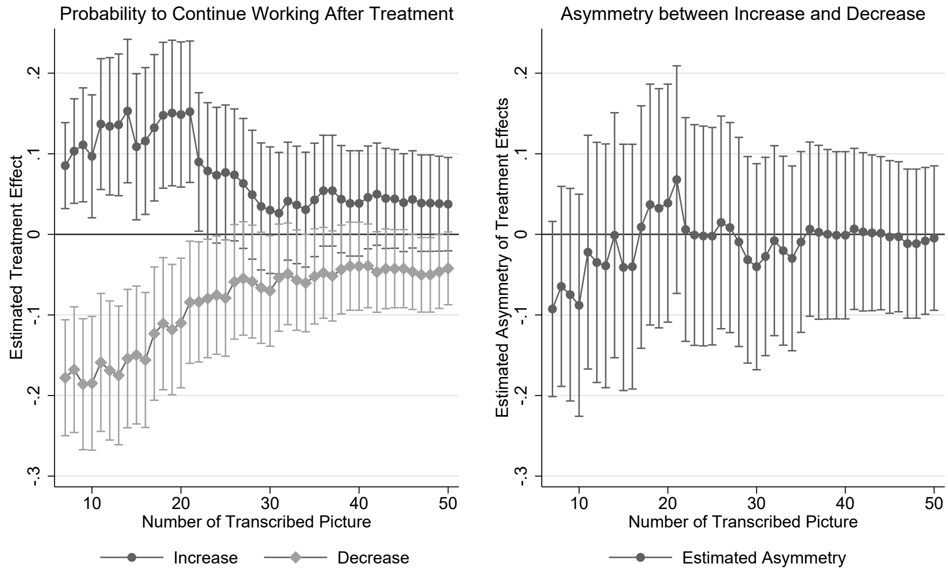
Notes: OLS regressions based on equation (3.1). Robust standard errors in parentheses (* significant at 10 %, ** significant at 5 %, *** significant at 1 %). The explanatory variables of interest are dummies indicating the *Wage Increase* and *Wage Decrease* group, respectively. The coefficients are relative to the omitted *Control* group. The outcome variables in columns (1) to (5) are: (1) *Extensive Margin* is the extensive margin measured as a dummy variable indicating whether a subject quit the task immediately after seeing the treatment notification. (2) *Pictures* is the number of images transcribed. (3) *Total Time* is the time (in minutes) that a subject totally spent on working on the labor task. (4) *Mean Time* is the time (in minutes) that a subject spent working on one image. (5) *Accuracy* is the share of characters that is transcribed correctly. N is the number of observations. R^2 denotes the R-squared. $p(Inc = -Dec)$ is the p-value from a t-test testing whether the coefficients for the *Increase* and *Decrease* group add up to zero. *Elast. Increase* and *Elast. Decrease* are the implied elasticities in the treatment groups that indicate how the respective outcome variable responds to the wage change, using the control group as the counterfactual (see equation (3.3) in Section 3.3.3).

group. Both effects are significantly different from zero for each period until the end of the experiment in $t = 50$. As expected, treatment effects at the extensive margin are especially strong immediately after the treatment and become less important the longer subjects continue working. The right panel of Figure 3.7 investigates the asymmetry of this effect, i.e., the sum of the estimated coefficients $\widehat{\gamma}^t_{Increase}$ and $\widehat{\gamma}^t_{Decrease}$ from equation (3.4). We find that the asymmetry in the extensive margin response mainly occurs immediately after the treatment notification and we cannot reject symmetry five or more periods after treatment notification.

3.4.3 Time Responses

This section describes results for time related outcome variables, again using the subsample of workers who saw the treatment notification. We show means along with medians to account for potential outliers in the time participants spent working.

Figure 3.7: Dynamics of Treatment Effects

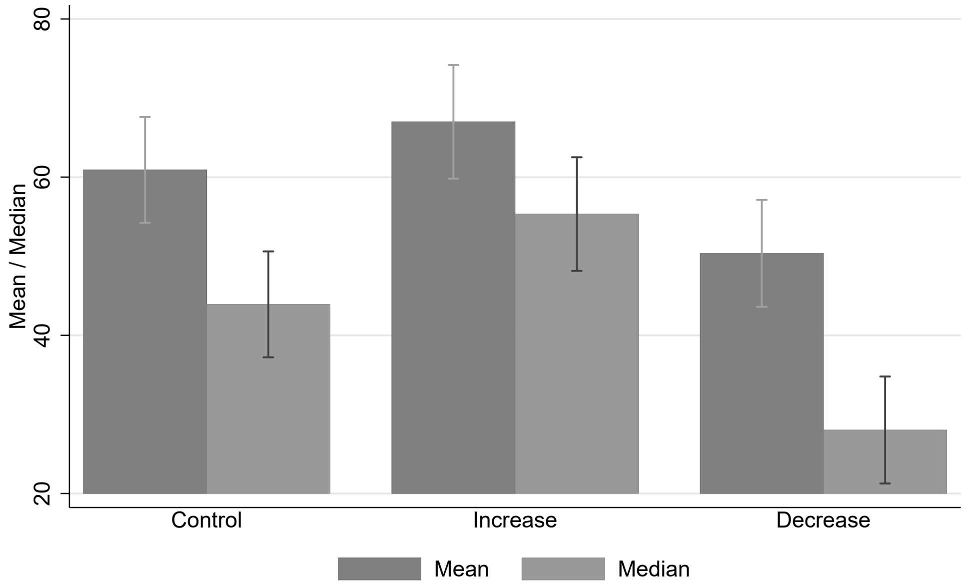


Notes: This figure shows the dynamics of the estimated treatment effects. The left panel shows coefficients that are based on regressions of the probability of working in a given period on a full interaction of an increase-group dummy and period dummies as well as a decrease-group dummy and period dummies (see equation (3.4) in the main body of the paper). The outcome variable is a dummy indicating if an individual works in the respective period (for example, this dummy takes value one in period ten if the respective individual transcribed the 10th picture). All effects are relative to the control group without wage change. The right panel shows the difference in left-panel regression coefficients between the *Wage Increase* and the *Wage Decrease* treatment group. All coefficients are shown along with 95 % confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the "wage increase" group and 257 subjects in the "wage decrease" group. All 720 subjects in the sample have transcribed at least six images.

Time Spent Working. Figure 3.8 shows that, on average, subjects in the control group spent about 61 minutes working on the labor task with a median of 44 minutes. Relative to the control group, workers who experienced a wage increase worked on the task for 6 additional minutes on average (11 minutes difference at the median) while those who experienced a wage decrease spent on average 11 fewer minutes (16 minutes at the median) working on the task. A non-parametric test shows that the treatment effect is statistically different from zero for the wage-decrease group, but not for the wage increase group. The non-parametric results are also reflected by the regressions; see Model III of Table 3.3. Again, these results indicate that labor supply and wages are positively related.

The differences indicate an asymmetric effect; the treatment effect is larger in the wage-decrease group than in the wage-increase group. This is also evident by the implied elasticities,

Figure 3.8: Total Time Worked by Treatment Group



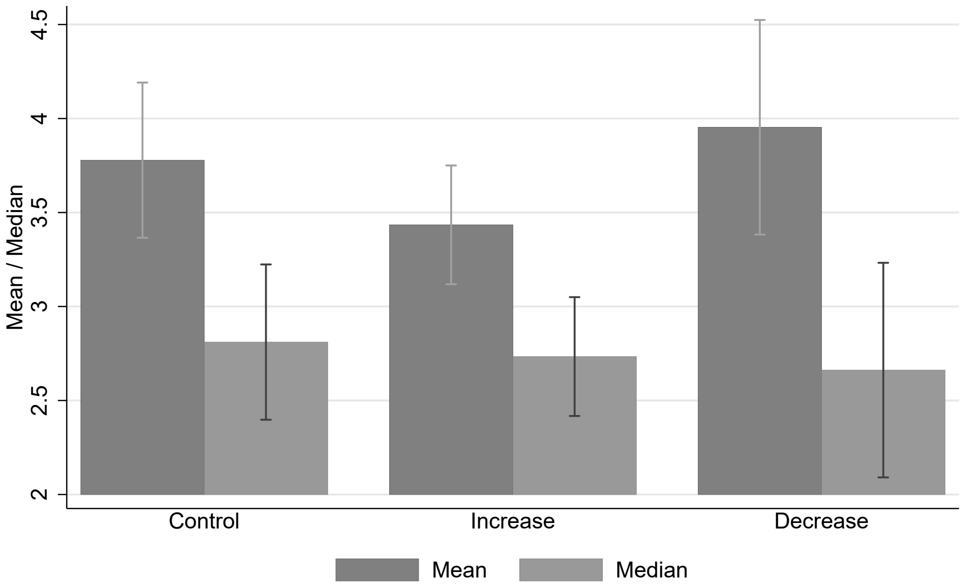
Notes: This figure depicts for each group the average and median time (in minutes) that subjects totally spent on working on the labor task, along with 95 % confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the "wage increase" group and 257 subjects in the "wage decrease" group. All 720 subjects in the sample have transcribed at least six images.

which are 0.50 in the increase group and 0.87 in the decrease group. However, we cannot reject the null that the difference between the treatment effects is zero. In other words, though the relative magnitude of the treatment effects is indicative of an asymmetric response, we cannot rule out symmetry in a statistical sense.

The effect on the total time spent working described above can be decomposed into two parts; the first due to the extensive margin response and the second due to the intensive margin response. We identify the contribution of the intensive margin response in Appendix Figure 3.A.2, which plots the mean and median of the total number of minutes worked conditional on *not* quitting right away after the treatment. The figure shows that, conditional on transcribing at least one picture after the treatment notification, workers in the control group spent an average of 68 minutes on the task (with a median of 50). Relative to the control group, workers in the wage-increase group worked for one additional minute on average (7 additional minutes at the median) while workers in the wage-decrease group spent on average 4 fewer minutes on the task (median 8). Subtracting these average intensive-time-margin treatment effects from the total treatment effects implies that the extensive margin response explains the overwhelming majority of the effect on time spent working on the task. In fact, the extensive margin response explains 83 % (= (6 − 1)/6) and 64 % (= (11 − 4)/11) of the time margin response in the wage increase and decrease groups, respectively.

Transcription Rate. The results for the transcription rate are shown in Figure 3.9. Workers on average spent 3.8, 3.4 and 3.9 minutes for one picture in the control, increase and decrease groups, respectively. The differences between groups are not statistically significant (also see Model IV in Table 3.3). We further separate this total effect into its intensive and extensive margin components and find that there is no statistically significant effect on either margin (see Appendix Figure 3.A.3 which reports the transcription rate conditional on completing at least one transcription after the treatment notification).

Figure 3.9: Average Time per Transcription by Treatment Group



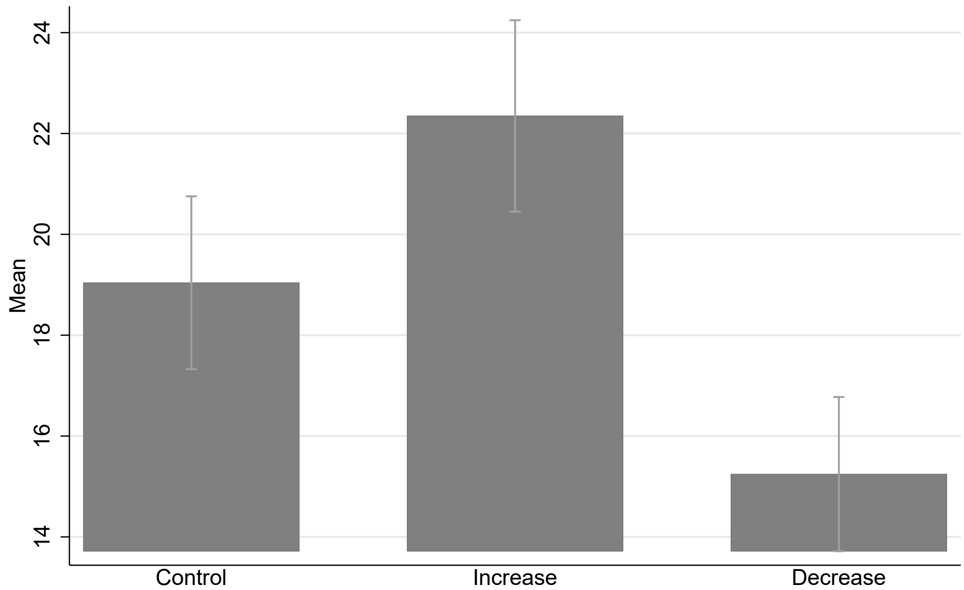
Notes: This figure depicts for each group the average and median time (in minutes) that subjects spent working on one image, along with 95 % confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the "wage increase" group and 257 subjects in the "wage decrease" group. All 720 subjects in the sample have transcribed at least six images.

3.4.4 Number and Quality of Transcriptions

This section describes the treatment effects on the number of transcribed pictures and the quality of transcription. All analyses are again based on the subsample of workers who saw the treatment notification.

Number of Transcribed Pictures. Figure 3.10 shows that the treatment variation clearly affected the number of transcribed pictures per worker. While the average worker transcribed 19.04 images in the control group, the average worker completed 22.35 and 15.25 pictures in the wage-increase and wage-decrease groups, respectively. All group-wise differences between groups are distinguishable from zero at the 1 %-level according to non-parametric rank-sum tests. The relationship between labor supply and wages is therefore again positive. These results are confirmed in Model II of Regression Table 3.3, which also shows that we cannot reject the null that the wage effect on total output is symmetric.

Figure 3.10: Number of Transcribed Pictures by Treatment Group



Notes: This figure depicts for each group the average number of images that subjects transcribed, along with 95 % confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the “wage increase” group and 257 subjects in the “wage decrease” group. All 720 subjects in the sample have transcribed at least six images.

As in Section 3.4.3, we decompose the total effect on number of transcribed pictures into its intensive and extensive margin components. We begin with the contribution of the intensive margin response by calculating the per-worker number of transcriptions for each group conditional on completing at least one picture after seeing the treatment information. These results, which are presented in Appendix Figure 3.A.4, show that output is higher when wages rise and lower when wages fall. While the non-parametric tests reveal that the difference between the control and increase group is statistically significant, the difference between control and decrease is not significant (p-value: 0.15). More importantly, the magnitude of these intensive-time-margin effects is not asymmetric in a statistical sense.

We next identify the contribution of the extensive margin response by subtracting the intensive-time-margin effect from the total effect. For example, the total treatment effect for the wage-increase group is 3.3 transcriptions. From Appendix Figure 3.A.4, we know that 2.14 of this effect is due to the intensive-time-margin response. Therefore, the balance of 1.17 ($= 3.31 - 2.14$) is due to the extensive margin response. A similar calculation for the wage-decrease group reveals that the contribution of the extensive margin is 2.19.

Quality of Transcriptions. Figure 3.A.5 in the Appendix depicts that the wage-rate changes did not have any effects on the quality of transcription. The differences are tiny and indistinguishable from zero, which confirms that workers in all groups paid careful attention to the task. This result is in line with the field experiment of Kube et al. (2013) who do not find any effects of wages on work quality either.

3.4.5 Robustness

Because the workers discussed our task on the mTurk forum, it is possible that our findings are driven by selection into the HIT. We explore this by performing the analyses separately on the sample of workers who worked on our task before it was discussed online and the sample of workers who worked on it afterwards. These results, which are presented in Appendix Figures 3.A.7–3.A.11, show no evidence that our results are driven by selection among workers who participated in the post-forum period. In addition, we regress each outcome variable on a dummy variable indicating whether the subject worked on the task before or after the forum post; we do not find any significant effects of working on the task after the forum post (results not reported).

3.5 Discussion of Results

In this section, we first discuss the potential economic explanations behind our results, and then describe the policy implications and generalizability of our findings.

3.5.1 Mechanisms

Our results show that the extensive margin response to wage changes is strongly asymmetric. We also find evidence of an asymmetric intensive-time-margin response, but this effect is not statistically distinguishable from zero. Similarly, the wage-induced effect on number of transcribed images is marginally asymmetric, though not statistically significant. How can these results be rationalized? In the following, we discuss several channels that help to understand the economic mechanisms behind our results.

Standard Labor Supply. One possible explanation of our *extensive-margin* results is that they are driven by a rational response to the difference between reservation wages and the newly announced wage: A worker's decision to work or not is determined by the wage rate relative to the worker's reservation wage, and the worker chooses to work if the wage rate is greater than her reservation wage. Since participation in our experiment is voluntary, it is reasonable to assume that the reservation wage for our workers has a distribution that is bounded between \$ 0 and \$ 0.15. This raises the possibility that some workers have reservation wage between \$ 0.12 and \$ 0.15. If true, this would make the observed responses consistent with a rational calculus of reservation wages. In particular, we would expect all rational workers with reservation wage between \$ 0.12 and \$ 0.15 to quit the labor task when the wage rate decreases to \$ 0.12. In contrast, workers whose wages stay constant or increase keep working because the new wages are at least as large as their reservation wage. As a result, the response to the wage decrease is larger than to the wage increase.

There are two potential reasons why it is not immediately clear that our results can be explained by this story of reservation wages. First, previous studies by Horton and Chilton (2010) and Horton et al. (2011) find that mTurkers have a median reservation wage between \$ 1.12 and \$ 1.38 per hour,[22] which is substantially lower than the implied hourly wage of \$ 2.10 in our wage decrease group.[23] However, although the median hourly reservation wage is considerably lower than the hourly wage in our decrease group, it is likely that the number of workers with reservation wage between \$ 0.12 and \$ 0.15 is *not* zero, suggesting that we should see a response to the wage decrease at the extensive margin even in light of a very low average reservation wage.

Second, we observe a statistically significant extensive margin response in the wage-increase group, which, at first glance, appears inconsistent with the reservation-wage argument since every worker in this group would have been paid above her reservation wage from the beginning of the experiment. However, the reservation-wage story might explain this result if workers have imperfect information about the disutility of the labor task before they start working on

[22] Horton et al. (2011) estimate a median reservation wage of \$ 1.12 using data generated from an mTurk task that is similar to our task. This task required mTurk workers to transcribe paragraph-sized chunks of text that are written in Tagalog, a language of the Philippines. That is, as in our task, subjects are required to transcribe foreign language (workers in their task were not from the Philippines) and are paid per transcribed text.

[23] The implied hourly wage in the decrease group is calculated based on the observation that workers in the decrease group transcribe about 15.18 pictures per hour: $2.10 = 0.1 + 6 \times 0.15 + 9.18 \times 0.12$. Note that this hourly wage is a lower bound because our measure of the time it takes to transcribe one picture overstates the actual time per picture; see Section 3.3.1.

our task. Because workers make their decision to start working on our task based solely on our description on the mTurk website, they are only able to form an expectation regarding the disutility of the labor task. Once workers transcribe the first batch of six pictures, they are able to update their estimate of the costs of working and the decision to continue working after this first batch of transcriptions is based on this updated estimate. Some subjects may have underestimated the disutility of working on the task and will quit the task after the first pictures even in the absence of any wage changes. This is consistent with our observation of positive quit-rates in the control group.[24] This mechanism additionally suggests that the share of workers who quit after a wage decrease is larger than in the control group without wage changes. It also suggests that the share of workers who quit in the wage-increase group is smaller than in the control group, but potentially still positive. As a result, this argument of updated beliefs about the disutility of the labor task, along with a non-zero number of workers whose reservation is between \$ 0.12 and \$ 0.15, provides a rationale for an asymmetric labor supply response to wage increases and decreases, which is consistent with a standard labor-supply model with reservation wages.

So what about the *intensive-margin* results? Could these results be due to the standard model? Notice that the intensive margin response is based on the difference between the marginal disutility of transcription and the wage rate. Assuming the disutility of transcription is increasing in the number of transcriptions, we would expect workers in the wage-increase group to work longer and faster, relative to the control group. On the other hand, because the wage-decrease group faces a lower wage than the control group, we would expect workers in this group to spend less time working and to do so at a slower rate. This is exactly what we find, implying that we can confirm a positive relation between labor supply and wages. We further find indications that the economic magnitude of these responses is asymmetric; e.g., the intensive-margin treatment effect for the time spent working in the wage-decrease group is four times that in the wage-increase group. The standard neoclassical labor supply model may yield such asymmetric labor supply responses if the labor supply function has a particular shape where an increase triggers a smaller response than a decrease.[25] Even then, asymmetry would only arise for non-marginal wage changes. Although our absolute change in the wage rate of \$ 0.03 is small, the relative change is 20 % and therefore unlikely to be perceived as marginal. This implies that our findings regarding the intensive margin are also consistent with a standard model of labor supply.

[24] Note that another reason for quits in the control group is that we intentionally nudge workers to quit after six pictures; see the notification that we display to workers in all groups after the first six transcribed pictures.

[25] For example, a CRRA utility function $\frac{1}{1-\gamma}(wL)^{1-\gamma}$ with linear costs aL yields asymmetric labor supply responses for non-marginal wage changes.

Loss Aversion. Our findings can also be rationalized by a model of loss aversion where the reference wage is equal to the expected wage of \$ 0.15 and the labor supply curve is kinked at this reference wage (a model of this type has for example been put forward by Ahrens et al., 2014; see Section 3.B in the Appendix).[26] In such a model, workers are loss averse with reference-dependent utility functions; workers' labor-supply functions are kinked at at a reference wage and have steeper slopes in the gain domain than in the loss domain. As a result, workers are less willing to supply an additional unit of labor when the wage is above the reference wage than when it is below, and a wage decrease is predicted to have a larger labor-supply effect than a wage increase of equal magnitude.

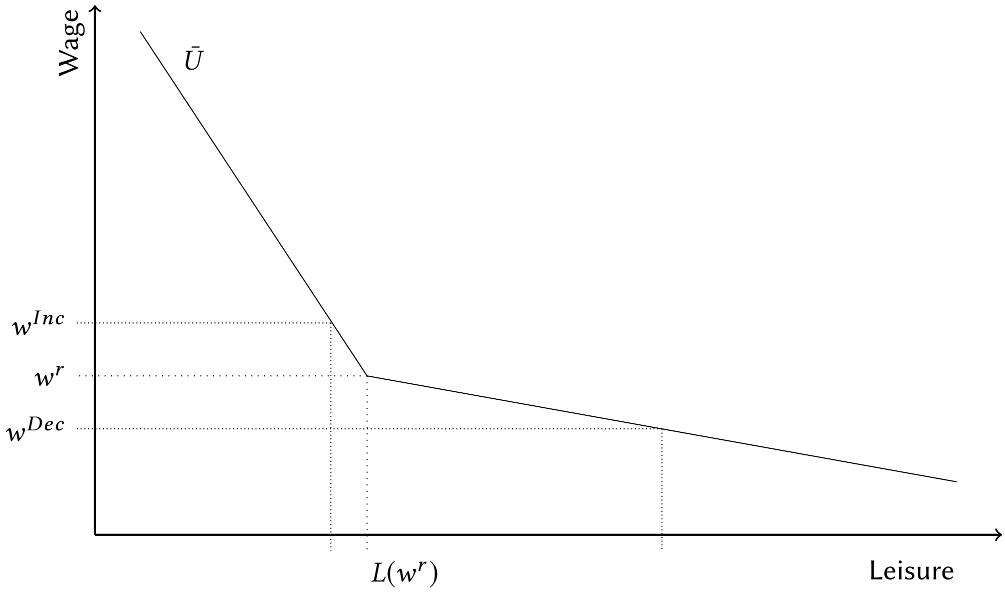
The main insight from this potential mechanism is sketched in Figure 3.11, which relates leisure and wages. A worker who is located at the reference wage, denoted w^r, will respond differently to wage increases and decreases of equal magnitude. In particular, a worker at the reference point weights wage decreases more heavily than wage increases. As a result, she will respond more strongly to a wage decrease (by working less) than an equally sized wage increase (to which she will respond through more labor supply). This result implies that labor supply elasticities identified from wage increases are predicted to be smaller than labor supply elasticities identified from wage decreases. Our findings are consistent with this prediction.

Reciprocity. Another potential explanation of our findings is reciprocity; workers interpret the wage changes as punishment or reward, and respond accordingly. Workers who receive a wage decrease feel punished and thus lower their labor supply in an effort to punish the employer, while workers who receive a wage increase feel rewarded for their effort and thus work harder to return the favor to their employer. To the extent that the degree of induced reciprocity is asymmetric, this explanation is potentially consistent with our findings. Although we have no way of ruling out this motivation behind our results, we argue that this is an unlikely explanation based on our experimental design. Recall that subjects are paid for each completed transcription and not per unit of time. This implies that workers in our experiment have little scope for punishing the employer through shirking. Reducing the number of transcription implies that a worker punishes herself in the form of lower pay-off, and potentially lower performance rating, which affects her prospects of being allowed to work on other mTurk tasks.[27] One strategy to punish the employer without incurring a cost is to continue to work

[26] It is reasonable to assume that workers in our field experiment expect the per-unit wage to remain constant at \$ 0.15. The literature typically finds that reference points depend on expectations (e.g., Kőszegi and Rabin, 2006, Abeler et al., 2011, Ahrens et al., 2014), suggesting that \$ 0.15 would constitute the reference wage if a model of loss aversion was applied to our context.

[27] Workers on mTurk receive ratings for each task they complete. Employers often use workers' performance rating to screen out low performers from their tasks. Therefore, a worker who decides to punish us because their wage has been reduced, runs the risk of limiting the number of tasks she will qualify to work on in the future.

Figure 3.11: Labor Supply under Loss Aversion



Notes: This figure displays the relationship between leisure and wages under loss aversion. The indifference curve is kinked at the reference wage w^r. Individuals who currently face the reference wage respond stronger to a wage decrease (by supplying less labor) than to a wage increase of equal magnitude (by supplying more labor).

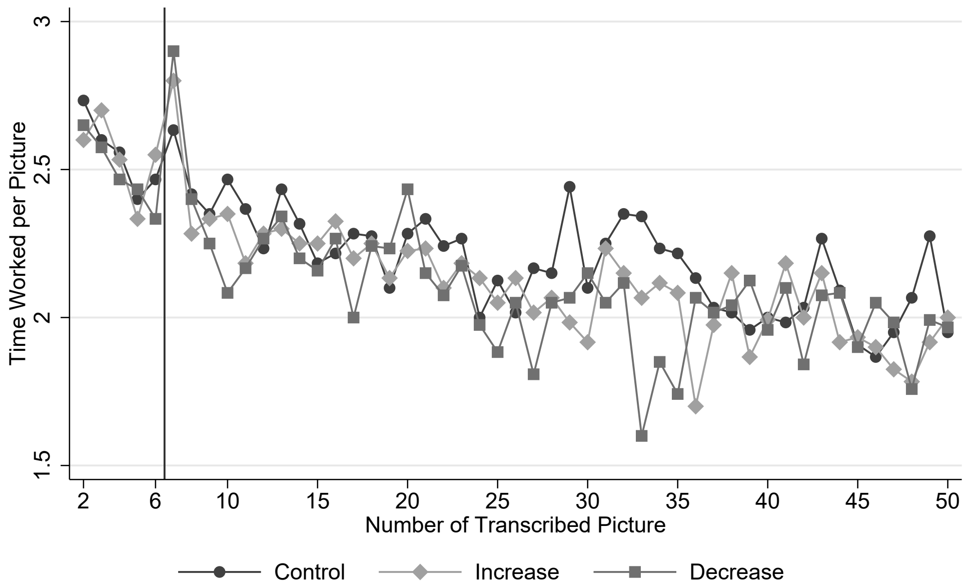
hard, but submit transcriptions that are of low enough quality to be mostly useless to the employer, but high enough quality to avoid a negative performance review. Because we have the transcriptions and the actual texts, we can check to see if workers used this strategy; there is no evidence that they did (see Section 3.4.4).

Similarly, as opposed to settings where workers are paid per hour, transcribing more pictures is not a reward for the employer in our experiment since this increases the costs to the employer. Workers are also likely to know that employers can easily recruit other workers to transcribe pictures and that employers therefore do not face the risk that pictures remain untranscribed.

Treatment-Induced Heterogeneous Skill Composition. One could think that the skill composition (i.e., the ability to transcribe pictures) of workers in the three experimental groups is affected by the treatment variations. For example, it might be plausible that the least productive workers drop out of the labor task once they are faced with a wage decrease, whereas similar unskilled workers stay after a wage increase. This would then imply that the share of unskilled workers in the decrease group would be smaller after treatment notification than in the increase group. As a result, differences across groups might simply be a result of heterogeneous attrition due to the treatments.

Using all individuals in our work task (thus not only those who saw the treatment), Figure 3.12 and Appendix Figure 3.A.6 show that transcription skills do not affect drop-out rates. Figure 3.12 shows the median time per transcribed picture over the course of all periods.[28] We use median time per transcribed picture as a measure of transcription skills. The figure shows that average transcription skills in the three experimental groups are not affected by the treatment notification; the lines evolve similarly over the periods in all three groups. Figure 3.A.6 in the Appendix shows the estimates of a regression of a 'Last-picture-transcribed' dummy (i.e., a dummy indicating if the respective period is the last period of the respective individual) on a full interaction of period dummies and the transcription time in the respective period. The figure shows that transcription skills (as measured by transcription time) do not determine exit rates. In other words, being talented or not in transcribing pictures does not predict if a participant drops out of the labor task.

Figure 3.12: Time Worked per Picture Over Periods, by Experimental Group



Notes: This figure depicts the median time it takes to transcribe a picture in each period by experimental group. For example, the value in period ten indicates the median time it took subjects in a given experimental group to transcribe the 10th picture. Treatment occurred after period six. The sample includes all participants who started working on the task. Number of observations is 1,152.

[28] For example, in period 10 the figure shows the median time that it took workers in the experimental groups to transcribe the 10th picture.

3.5.2 Implications

The existing labor-supply literature often identifies labor supply elasticities by exploiting panel data comprised of both wage increases and decreases. This approach makes sense in the context of *marginal* wage changes where the elasticity is shown to be symmetric. However, this approach becomes problematic when on considers that wage changes are generally non-marginal. The reason is that both the standard and behaviorally-inspired models can be used to show that labor supply responses to non-marginal wage changes need not be symmetric. Consistent with the theoretical finding of Ahrens et al. (2014) and the empirical results of Kube et al. (2013), we show that labor supply responds asymmetrically to non-marginal wage changes.

Our findings suggest that ignoring the direction of wage changes when estimating labor supply elasticities leads to biased own-wage labor-supply elasticities; elasticities are overestimated when wages rise and underestimated when wages fall. Importantly, and as an addition to the previous literature, we find that the asymmetry of labor supply w.r.t. wages is more pronounced on the extensive margin relative to the intensive margin. This refinement of the asymmetry is especially important since it is generally accepted that labor supply elasticities are mainly determined by the extensive margin response (Blundell and MaCurdy, 1999, Meghir and Phillips, 2010, Bargain et al., 2014). Our findings are also practically useful, since labor supply elasticities play an important role in quantifying the economic impacts of policy changes that affect wages; e.g., minimum wage polices.

Additionally, our results highlight one potential reason for the downward rigidity in wages. In particular, it is widely observed that nominal wages tend to move in one direction only. Prominent explanations for downward nominal rigidities include institutions such as minimum wages and collective bargaining. Recent evidence by Kaur (2018), however, shows that wages are downward rigid even in the absence of such institutions. Among the many potential explanations for this observed rigidity are the potentially significant effects on productivity and labor supply. However, the scarcity of nominal wage cuts makes it challenging to determine if the labor supply and production impacts of nominal wage cuts are indeed large and negative. Our study adds to the small literature that have explored this question. As mentioned before, we find large negative extensive margin responses, which suggest that nominal wage cuts could be potentially damaging for firms. This is one reason behind the reluctance of firms to reduce wages.

A policy area where our findings are likely to be particularly useful is taxation. First, tax reforms generally result in either tax increases or tax decreases, which translate into changing after-tax wages. In fact, upward and downward changes are more prominent for tax rates than for wages. We know that the tax elasticity of labor supply is generally larger than the wage elasticity; e.g., due to tax aversion (Kessler and Norton, 2016). This suggests that the labor-supply asymmetry with respect to tax-rate changes is likely to be more pronounced than what our findings for labor supply responses to wage changes suggest. This makes the distinction between rate increases and decreases particularly important in the context of tax rate. Second, our findings also raise questions about the elasticity of taxable income (ETI) which plays a crucial role in our understanding of the efficiency costs of taxation and which is often estimated based on panel-data exploiting multiple variations in tax rates (e.g., Saez et al., 2012). In particular, our results suggest that failure to distinguish between ETI estimated with tax rate increases and ETI estimated with tax rate decreases is likely to lead to an underestimation of the efficiency cost of tax rate increases. This problem is likely to be even more important than with wage changes since tax rates generally move freely in both directions. Of course, the labor supply response to wage changes is not necessarily identical to the response to tax rate changes. Therefore, we are cautious in generalizing our results to the case of tax rate changes.[29]

3.5.3 Generalizability

The results described above are obtained using an experimental design in a large real-world labor market. Importantly, workers did not know they participated in an experiment and thus behaved as they would in their natural occurring environment. At least two aspects of our empirical results suggest that workers took the task seriously and supplied labor in a “plausible” way. First, the overall quality is very high with an accuracy rate of more than 97 % in all three groups. Second, we find evidence for an upward sloping labor supply curve – that is, workers work more as wages increase and less when wages go down –, which is consistent with the empirical labor supply literature (e.g., Keane, 2011, Bargain et al., 2014). These two points also indicate that labor-supply behavior in our online task has some implications for labor-supply behavior in other, more traditional labor markets.

Due to randomization, our experimental design also guarantees internal validity. Though our findings are based on an actual real-world labor market, we are careful not to generalize our results to all types of labor markets. Nonetheless, we argue that the findings are applicable to labor markets with piece rate, flexibility and multiple outside options. One example of such labor markets is on-line crowd-sourcing labor markets, which are becoming increasingly common in the current technological age.[30] A common feature of these labor markets is that workers tend to work for relatively low wages and have extremely high levels of labor supply flexibility. Because the labor supply effects are predominately on the extensive margin, we argue that the

[29] We are not aware of any evidence on the asymmetric effects of tax increases and decreases on labor supply or taxable income. Benzarti et al. (2017) show that increases in Value Added Taxes (VAT) have a larger effect on prices than VAT reductions.

[30] See `https://sites.google.com/site/johnjosephhorton/miscellany/online-labor-markets` for a list of crowd-sourcing online labor markets.

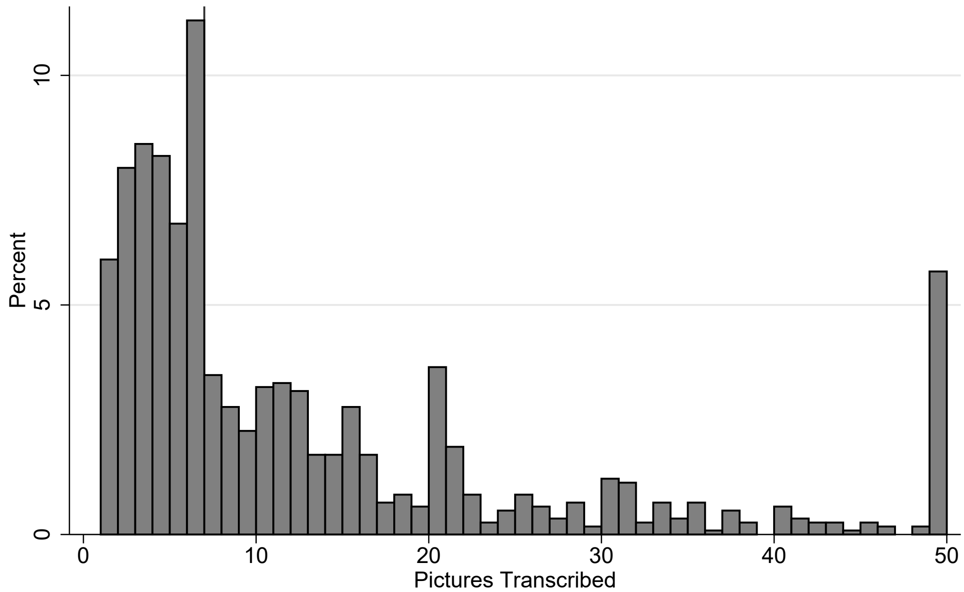
results are also likely to be equally applicable to traditional labor markets where workers face greater restrictions on labor hours.

3.6 Conclusion

We estimate the effect of (non-marginal) wage changes on labor supply using data generated in a field experiment on Amazon's Mechanical Turk. Our findings show that the labor-supply curve for workers on mTurk is upward sloping; the relationship between wages and labor supply is positive both for the case of wage increases and wage decreases. We further find strong evidence of an asymmetric response on the extensive margin; the extensive-margin response for wage decreases is twice as large as for equally sized increases. The magnitude of the intensive-time margin responses is also indicative of an asymmetric response, but we cannot rule out symmetry in a statistical sense. These results are consistent with a standard labor-supply model as well as reference-dependent preferences. Importantly, they are not driven by treatment-induced heterogeneous skill compositions.

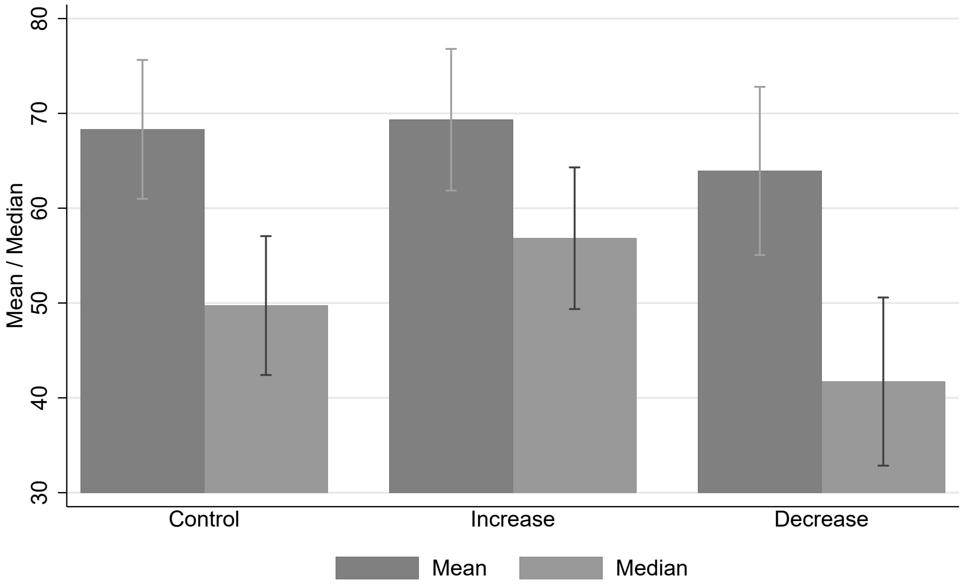
Appendix 3.A Additional Figures

Figure 3.A.1: Histogram of Transcribed Pictures



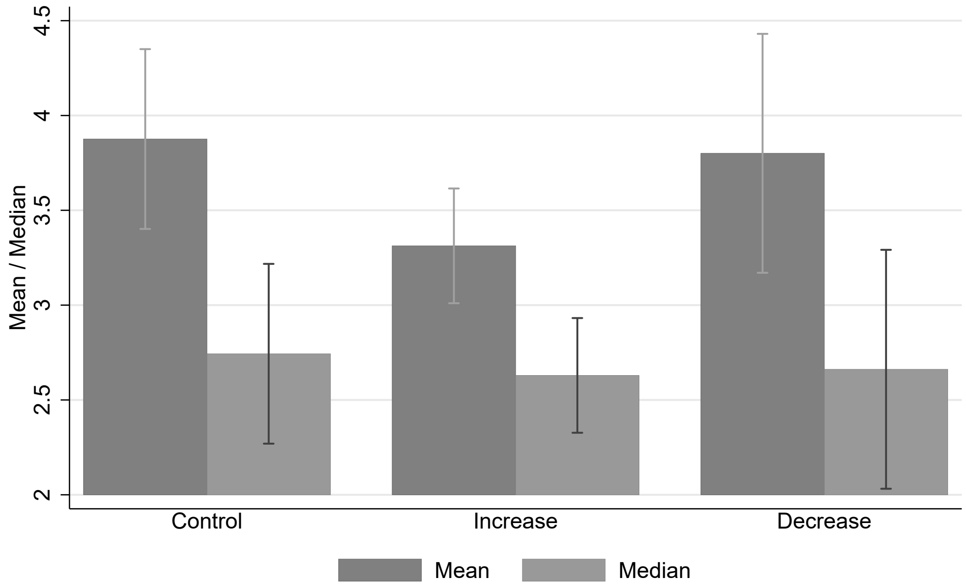
Notes: This figure plots the histogram of pictures described for all workers who worked on the task. The number of observations is 1,152. Subjects saw the treatment notification after transcribing six pictures (indicated by the vertical line).

Figure 3.A.2: Total Time Worked by Treatment Group – Intensive Margin



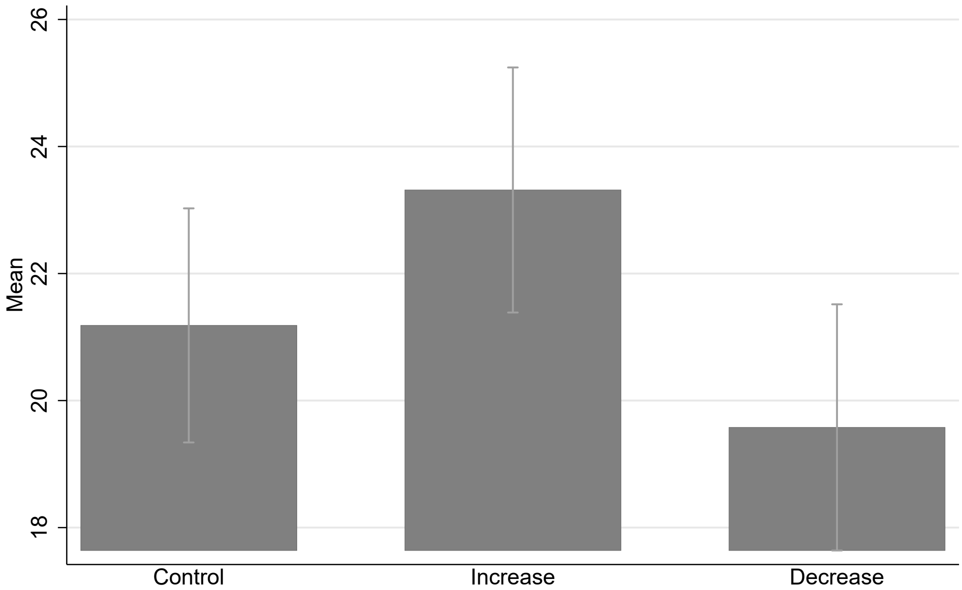
Notes: This figure depicts for each group the average and median time (in minutes) that subjects totally spent on working on the labor task, along with 95 % confidence bands. The underlying sample is restricted to subjects who did *not* quit the labor task immediately after seeing the treatment notification (i.e., restricted to subjects who have transcribed at least seven images). The number of observations is 591 with 213 subjects in the control group, 203 subjects in the "wage increase" group and 175 subjects in the "wage decrease" group. All 591 subjects in the sample have transcribed at least seven images.

Figure 3.A.3: Average Time per Transcription by Treatment Group – Intensive Margin



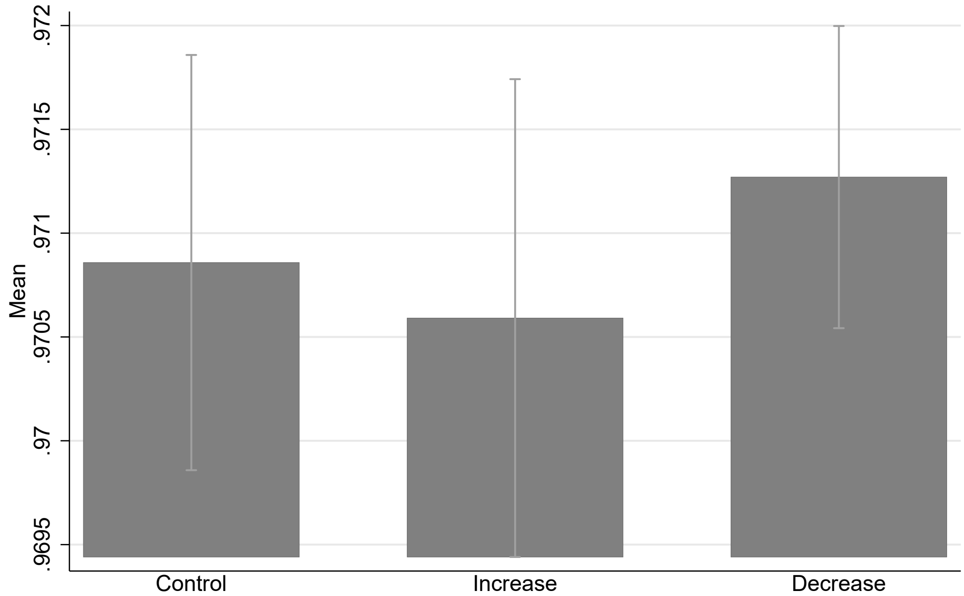
Notes: This figure depicts for each group the average and median time (in minutes) that subjects spent working on one image, along with 95 % confidence bands. The underlying sample is restricted to subjects who did *not* quit the labor task immediately after seeing the treatment notification (i.e., restricted to subjects who have transcribed at least seven images). The number of observations is 591 with 213 subjects in the control group, 203 subjects in the "wage increase" group and 175 subjects in the "wage decrease" group. All 591 subjects in the sample have transcribed at least seven images.

Figure 3.A.4: Number of Transcribed Pictures for Workers Who Completed at Least One Picture After the Treatment Notification



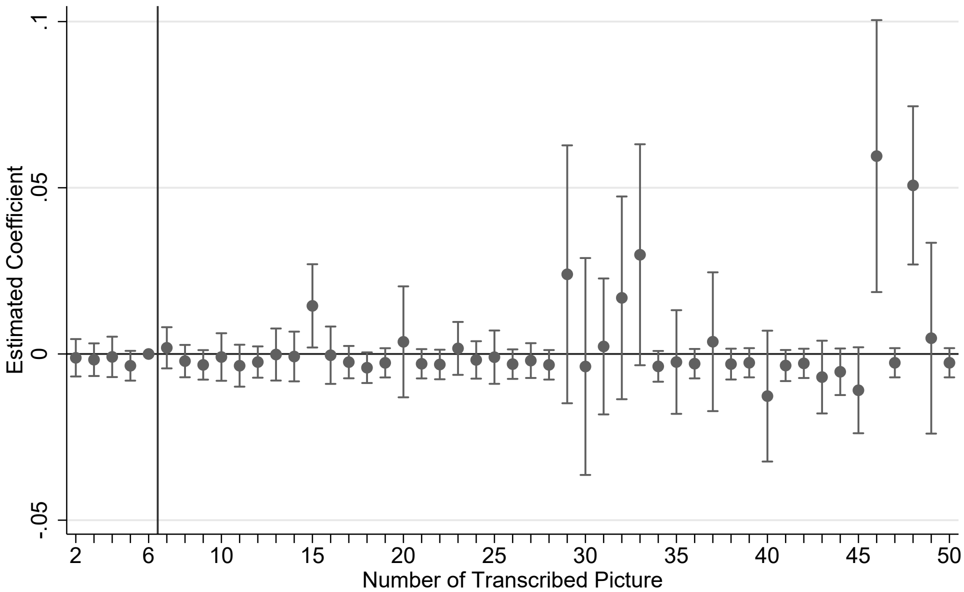
Notes: This figure depicts for each group the average number of images that subjects transcribed, along with 95 % confidence bands. The underlying sample is restricted to subjects who did *not* quit the labor task immediately after seeing the treatment notification (i.e., restricted to subjects who have transcribed at least seven images). The number of observations is 591 with 213 subjects in the control group, 203 subjects in the “wage increase” group and 175 subjects in the “wage decrease” group. All 591 subjects in the sample have transcribed at least seven images.

Figure 3.A.5: Accuracy by Treatment Group



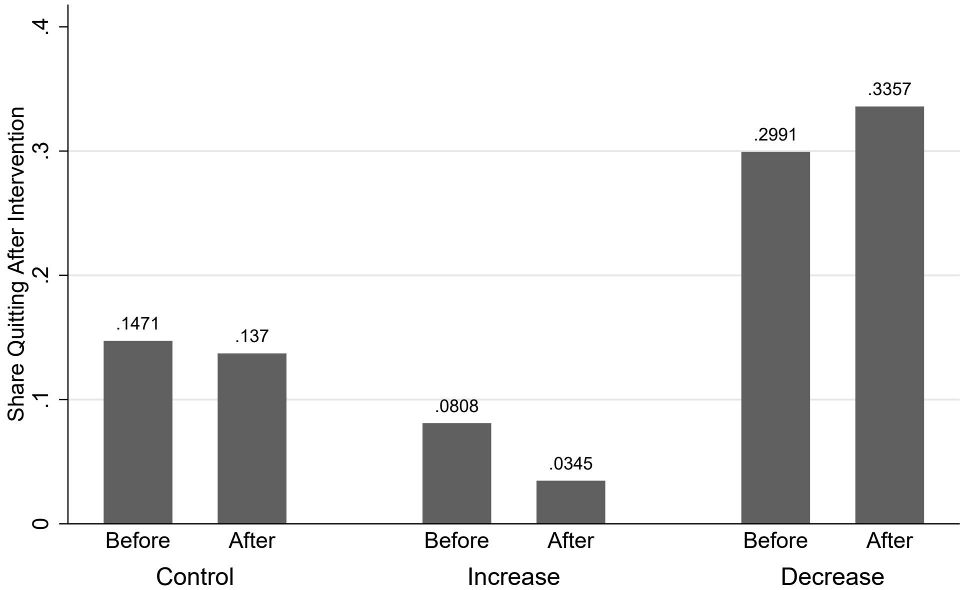
Notes: This figure depicts for each group the average transcription accuracy, i.e., the average share of characters in each image that is transcribed correctly, along with 95 % confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the “wage increase” group and 257 subjects in the “wage decrease” group. All 720 subjects in the sample have transcribed at least six images.

Figure 3.A.6: Effect of Performance on Probability to Quit



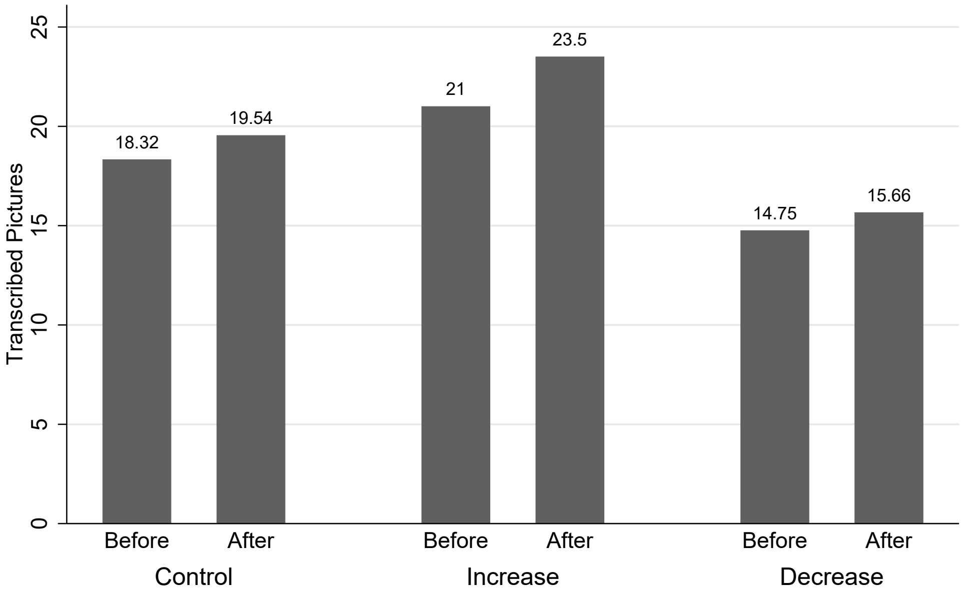
Notes: This figure depicts coefficients which are based on a regression of a “Last-picture-transcribed” dummy on a full interaction of period dummies and the time used to transcribe the respective picture. “Last-picture-transcribed” dummy is one if the picture in this respective round is the last transcribed picture of the respective participant. Coefficients are shown along with 95 % confidence bars. Standard error clustered by individual. Treatment occurred after period six and we normalize the coefficients to the pre-treatment period. The sample includes all participants who started working on the task. Number of observations is 1,152.

Figure 3.A.7: Extensive Margin – Before vs. After Forum Post



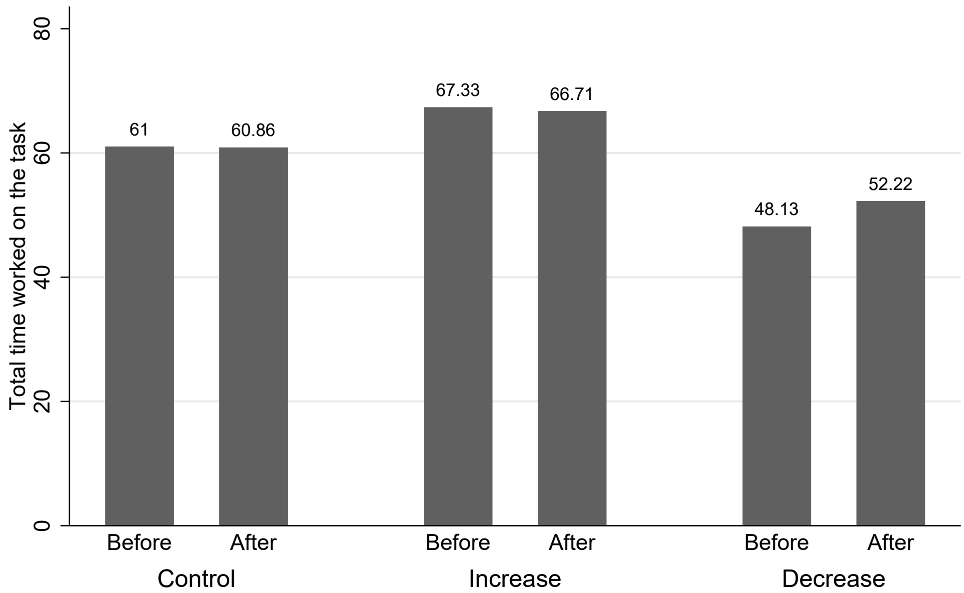
Notes: This figure depicts the share of subjects in each group who quit the labor task immediately after seeing the treatment notification (i.e., share of subjects who transcribed six pictures but not a seventh one). *Before* and *After* indicate whether the observation was sampled before or after the task was discussed online (see Section 3.3.2). The number of observations sampled before the forum post is 318 with 102 subjects in the control group, 99 subjects in the "wage increase" group and 117 subjects in the "wage decrease" group. The number of observations sampled after the forum post is 402 with 146 subjects in the control group, 116 subjects in the "wage increase" group and 140 subjects in the "wage decrease" group. All subjects in the sample have transcribed at least six images.

Figure 3.A.8: Number of Transcribed Pictures – Before vs. After Forum Post



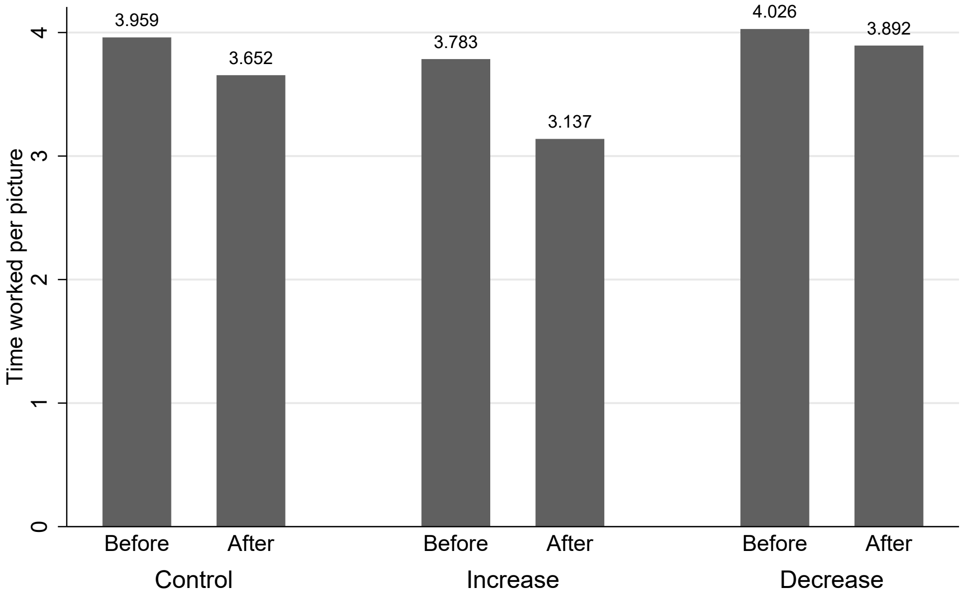
Notes: This figure depicts for each group the average number of images that subjects transcribed. *Before* and *After* indicate whether the observation was sampled before or after the task was discussed online (see Section 3.3.2). The number of observations sampled before the forum post is 318 with 102 subjects in the control group, 99 subjects in the "wage increase" group and 117 subjects in the "wage decrease" group. The number of observations sampled after the forum post is 402 with 146 subjects in the control group, 116 subjects in the "wage increase" group and 140 subjects in the "wage decrease" group. All subjects in the sample have transcribed at least six images.

Figure 3.A.9: Total Time Worked – Before vs. After Forum Post



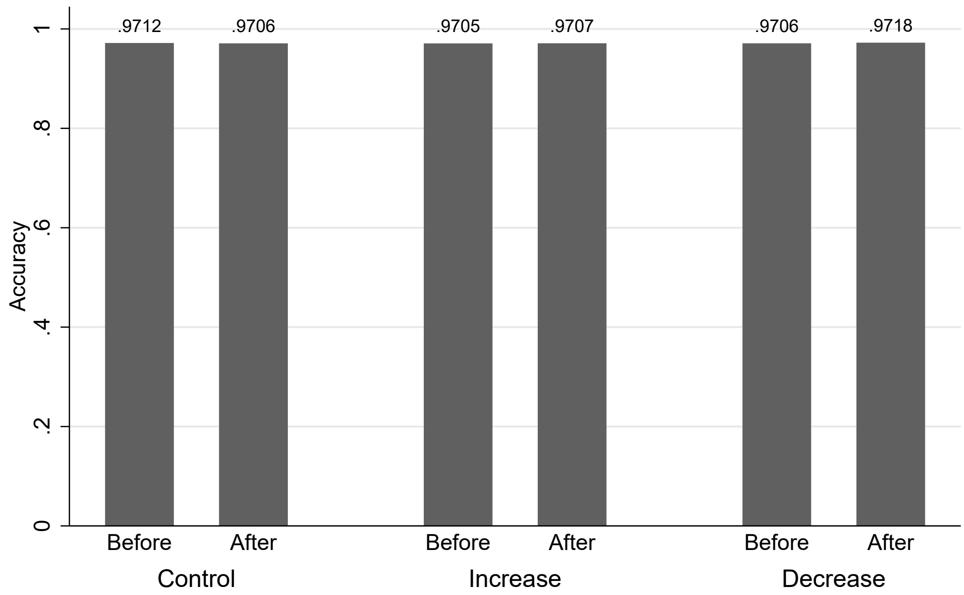
Notes: This figure depicts for each group the average time (in minutes) that subjects totally spent on working on the labor task. *Before* and *After* indicate whether the observation was sampled before or after the task was discussed online (see Section 3.3.2). The number of observations sampled before the forum post is 318 with 102 subjects in the control group, 99 subjects in the "wage increase" group and 117 subjects in the "wage decrease" group. The number of observations sampled after the forum post is 402 with 146 subjects in the control group, 116 subjects in the "wage increase" group and 140 subjects in the "wage decrease" group. All subjects in the sample have transcribed at least six images.

Figure 3.A.10: Average Time per HIT – Before vs. After Forum Post



Notes: This figure depicts for each group the average time (in minutes) that subjects spent working on one image. *Before* and *After* indicate whether the observation was sampled before or after the task was discussed online (see Section 3.3.2). The number of observations sampled before the forum post is 318 with 102 subjects in the control group, 99 subjects in the “wage increase” group and 117 subjects in the “wage decrease” group. The number of observations sampled after the forum post is 402 with 146 subjects in the control group, 116 subjects in the “wage increase” group and 140 subjects in the “wage decrease” group. All subjects in the sample have transcribed at least six images.

Figure 3.A.11: Accuracy – Before vs. After Forum Post



Notes: This figure depicts for each group the average transcription accuracy, i.e., the average share of characters in each image that is transcribed correctly. *Before* and *After* indicate whether the observation was sampled before or after the task was discussed online (see Section 3.3.2). The number of observations sampled before the forum post is 318 with 102 subjects in the control group, 99 subjects in the "wage increase" group and 117 subjects in the "wage decrease" group. The number of observations sampled after the forum post is 402 with 146 subjects in the control group, 116 subjects in the "wage increase" group and 140 subjects in the "wage decrease" group. All subjects in the sample have transcribed at least six images.

Appendix 3.B A Model of Labor Supply under Loss Aversion

This section presents a theoretical framework that allows us to predict the impact of wage increases and decreases on labor-supply. The outline is informed by Ahrens et al. (2014) who incorporate loss aversion into a standard labor-supply model.

Ahrens et al. (2014) develop a model where workers with reference-dependent preferences maximize the following utility function:

$$U(C, L) = U^C(C) - \theta_i \frac{L^{\vartheta_i}}{\vartheta_i},$$

where C is consumption, L is labor supply (hours worked or effort), and θ_i is a parameter to ensure preference continuity at the reference wage. $U^C(C)$ is utility from consumption and the term $\frac{L^{\vartheta_i}}{\vartheta_i}$ indicates disutility from working. ϑ_i is a measure of loss aversion, which is characterized by the following piece-wise function:

$$\vartheta_i = \begin{cases} \vartheta_1 & \text{if } w > w^r \\ \vartheta_2 & \text{if } w < w^r. \end{cases}$$

In this equation, w is the current wage (per unit of L supplied) and w^r is the reference wage.[31] If w is above the reference wage, the worker is in the so-called gain domain, and if w is below the reference wage, she is in the loss domain. A subject is loss averse if $\vartheta_1 > \vartheta_2$, implying that the marginal utility loss from working is higher in the gain domain than in the loss domain. This means that workers are less willing to supply an additional unit of labor when the wage is above the reference wage than when it is below. Maximizing with respect to the budget constraint $C = wL$ gives the following kinked labor-supply curve:[32]

$$L = \begin{cases} (\frac{w}{\theta_1})^{\frac{1}{\vartheta_1 - 1}} & \text{if } w > w^r \\ (\frac{w}{\theta_2})^{\frac{1}{\vartheta_2 - 1}} & \text{if } w < w^r \end{cases}$$

Because of loss aversion with respect to the reference wage w^r (and hence $\vartheta_1 > \vartheta_2$), we get that $\frac{1}{\vartheta_1 - 1} < \frac{1}{\vartheta_2 - 1}$. This implies that subjects whose current wage is the reference wage w^r are more responsive to wage decreases than to wage increases.[33]

[31] It is plausible to argue that \$ 0.15 constitutes the reference wage w^r in our set-up (see Section 3.5.1).

[32] We only discuss the main implications of the model here since Ahrens et al. (2014) has all of the derivations.

[33] We assume an upward sloping labor supply curve where the substitution effect dominates the income effect. That is, subjects work more when wages go up and they work less when wages fall. This assumption is also supported by our empirical findings.

Chapter 4

The Sensitivity of Structural Labor Supply Models to Modeling Assumptions*

4.1 Introduction

Knowing the size of labor supply responses to wage or policy changes has important implications for welfare analysis (Eissa et al., 2008) and optimal taxation (Diamond and Saez, 2011, Immervoll et al., 2011). Despite a long and comprehensive empirical literature on labor supply behavior, there is still substantial variation in the estimated elasticities (see, e.g., Heckman, 1993, Evers et al., 2008, Chetty et al., 2011, Keane and Rogerson, 2012). Potential reasons include differences in preferences, norms, and institutions across countries and over time. But even for the same country, the same period, and the same estimation approach there is still considerable heterogeneity in individuals' estimated responsiveness to wages (Bargain and Peichl, 2016). One explanation for these remaining differences is the use of different and/or wrongly specified empirical models.

In this paper, we aim to investigate this channel by thoroughly scrutinizing state-of-the-art micro labor supply models and their functioning.[1] Structural models are repeatedly criticized for the large number of assumptions and the even larger number of possible combinations of these assumptions (Keane, 2010, Manski, 2014). We test whether the numerous modeling choices actually affect estimated elasticities. More specifically, we check the internal validity of such models by running controlled experiments: we set up and estimate 3,456 different models, each representing a different (plausible) combination of commonly made assumptions. We use two different micro data sets – one for Germany and one for the US – and estimate these different models for five distinct population groups, leading to 17,280 maximum likelihood

* This chapter is based on a revised version of M. Löffler, A. Peichl, and S. Siegloch (2014b). "Structural Labor Supply Models and Wage Exogeneity". *IZA Discussion Paper* 8281.

[1] We focus on structural labor supply models which can be used for policy simulations. In addition, several reduced-form approaches are used in the literature to estimate labor supply responses (see Chetty et al., 2011, and Bargain and Peichl, 2016, for recent surveys).

estimations for each data set. Based on the estimation results, we gather insights into how robust the statistical fit of the models and the estimated labor supply elasticities are with respect to the underlying assumptions.

The modeling assumptions can be categorized in three broad areas. First, researchers need to specify the utility function. This concerns the functional form, its flexibility with respect to observed and unobserved preference heterogeneity, and the inclusion of stigma costs from welfare participation. Second, there are different ways to construct the choice set and to model the availability of job or hours alternatives and fixed costs of working. The third area relates to the treatment of the underlying wage distribution, namely, the imputation of wages for non-workers, whether to use predicted wages for non-workers only or to impute wages for the full sample, and the handling of the wage prediction error when imputing wages. While the second issue has been surveyed in Aaberge et al. (2009), the literature is rather silent on the first and the last area, which is were this paper intends to break new ground. In particular, the treatment of wages has received hardly any discussion in many existing studies.

Our results show that the models' predictions are strongly driven by the treatment of wages in the estimation. For instance, the choice between predicting wage rates for non-workers with missing wage information only or for the full sample – both procedures are often used in the literature – may double the estimated labor supply elasticities, raising the average own-wage elasticity in our meta-analysis from 0.23 to 0.46. While the former option presumes that individuals optimize with respect to their current wage, the latter specification assumes that all individuals base their labor supply decision on expected wages as derived from the Mincerian wage equation. The handling of wage prediction errors is equally important. Using predicted wages for all individuals but ignoring the forecast error yields an average elasticity of 0.65 as opposed to 0.35 when accounting for the prediction error. In contrast, it turns out that other modeling choices hardly affect the estimated results. Elasticities are largely robust to the specification of the functional form of the utility function, the inclusion of observed or unobserved preference heterogeneity, as well as the modeling of hours restrictions or stigma costs of welfare participation.

We conclude that the attention of previous sensitivity analyses has been mainly concentrated on less important factors while the main driving forces have been neglected, i.e., the interactions between wages, working hours, and preferences. This finding is even more relevant given that most existing models (implicitly) assume exogeneity between the wage equation and the labor supply decision.[2] Our findings have important policy implications as labor supply elasticities

[2] Only little effort has been made thus far in the context of discrete choice labor supply models to overcome this assumption. Aaberge et al. (1995), Breunig et al. (2008), and Blundell and Shephard (2012) estimate preferences and wages simultaneously, in part also allowing for correlation.

are key parameters when evaluating or designing optimal tax benefit policies. For instance, Diamond and Saez (2011) use an elasticity of 0.25 to derive an optimal top marginal tax rate of 72.7 percent. However, an elasticity of 0.65, as often found when using alternative wage imputation procedures, reduces the optimal tax rate to 50.6 percent, bringing it closer to actually observed values.

Our analysis makes two important contributions to the literature on labor supply estimation. First, there is little evidence on the functioning of structural labor supply models in general. Moreover, if such studies exist, different models are not estimated on the same data set. Existing surveys and meta-analyses focus on either the principles of alternative estimation strategies (Blundell and MaCurdy, 1999, Evers et al., 2008) or cross-country comparisons of empirical findings (Bargain et al., 2014). Robustness checks in previous studies usually limit themselves to small deviations in one or only few of the numerous modeling assumptions. In that respect, we run a controlled meta-analysis, isolating the impact of the model assumptions on estimation outcomes. Second, our analysis points to a hitherto neglected factor that strongly influences the estimated labor elasticities: we show that the treatment of wages in labor supply estimations, which is rarely theoretically motivated nor subject to robustness checks, crucially affects the estimation results.

The remainder of this paper is organized as follows. Section 4.2 presents the modeling framework and a short overview of the existing literature. Section 4.3 provides information on the used data and the modeling of the tax and benefit system. In Section 4.4 we conduct our analysis of modeling assumptions and present the results. Section 4.5 concludes.

4.2 Model and Existing Literature

The use of structural discrete choice labor supply estimations has become a standard procedure in the empirical analysis of labor supply for both econometricians and policy makers (see, e.g., the overview in Bargain and Peichl, 2016). The first generation of labor supply models relied on the assumption that the household's utility is maximized over a continuous set of working hours – known as *Hausman approach* (see Hausman, 1981). This approach has been criticized for three reasons: (i) because the consistent estimation relies on rather restrictive *a priori* assumptions (see, e.g., MaCurdy et al., 1990, or Bloemen and Kapteyn, 2008, for details); (ii) the procedure has proven cumbersome when the budget set is non-convex, which will often be the case in presence of complicated tax and benefits systems in most countries; (iii) it has been shown that the estimated elasticities are very sensitive to the underlying wages (Ericson and Flood, 1997, Eklöf and Sacklén, 2000).

Partly motivated by these shortcomings, it has become increasingly popular to model the labor supply decision as the choice between a (finite) set of utility levels instead of deriving the marginal utility. Starting with the works by Aaberge, Dagsvik, and Strøm (1995), van Soest (1995), and Hoynes (1996), a wide range of different empirical specifications of these *discrete choice models* has been applied. For many institutional settings, the assumption of a discrete choice between different working hours or job offers may even be more plausible than assuming a continuous choice set (Dagsvik et al., 2014). Comparing different levels of utility avoids also the cumbersome maximization process of Hausman-type models. We focus our analysis on the discrete choice approach, given that it has become the standard procedure in the literature.

4.2.1 General Model

Structural labor supply estimations build on the assumption of the well-known neoclassical labor supply model that decision makers maximize their utility by choosing the optimal amount of working hours (or, more generally, the optimal job) subject to a budget constraint. Utility is defined as a function of consumption C_{nj}, leisure L_j, and idiosyncratic preferences for certain jobs, which we denote by ϵ_{nj}. Individual n faces the decision between a set of jobs $j \in J_n$ with working hours h_j and wages w_{nj}, including non-participation,[3] and maximizes her utility over job alternatives:

$$\max_{j \in J_n} U\left(C_{nj}, L_j, \epsilon_{nj}\right) = \max_{j \in J_n} U\left(f\left[w_{nj}h_j, I_n \middle| \boldsymbol{x_{nj}}\right], T - h_j, \epsilon_{nj}\right) \tag{4.1}$$

where leisure L_j is denoted as difference between the total time endowment T and working hours h_j. Consumption C_{nj} depends on working hours, the hourly wage rate w_{nj}, non-labor income I_n, household and job characteristics $\boldsymbol{x_{nj}}$, and the tax benefit system $f[\cdot]$. We assume a static model, which implies that consumption equals disposable income.

Individuals' true utility is only partly observable to the researcher while idiosyncratic components captured in ϵ_{nj} are latent. We rewrite the utility of individual n choosing job type j accordingly as:

$$U\left(C_{nj}, L_j, \epsilon_{nj} \middle| \boldsymbol{x_{nj}}, \boldsymbol{\beta_n}, \boldsymbol{\gamma_j}\right) = \varphi\left(C_{nj}, L_j \middle| \boldsymbol{x_{nj}}, \boldsymbol{\beta_n}, \boldsymbol{\gamma_j}\right) + \epsilon_{nj} \tag{4.2}$$

The first part $\varphi\left(C_{nj}, L_j \middle| \boldsymbol{x_{nj}}, \boldsymbol{\beta_n}, \boldsymbol{\gamma_j}\right)$ is determined by consumption and leisure, characteristics $\boldsymbol{x_{nj}}$, individual preferences $\boldsymbol{\beta_n}$, and labor market conditions $\boldsymbol{\gamma_j}$ that capture the availability of job type j. One may think of these labor market characteristics $\boldsymbol{\gamma_j}$ as measuring fixed costs of working, search costs for part-time jobs or rigidities regarding working hours, for example. The unobserved taste variation ϵ_{nj} is assumed to be i.i.d. and follow the extreme value type I

[3] We denote non-participation as job alternative $j = 0$ with $h_0 = 0$ and $w_{n0} = 0$.

distribution with cumulative distribution function $F(\epsilon) = \exp(-\exp(-\epsilon))$. McFadden (1974) has shown that the probability of individual n choosing a job of type i is subsequently given by:

$$\mathrm{P}\left(U_{ni} > U_{nj}, \forall j \neq i \middle| \boldsymbol{x_n}, \boldsymbol{\beta_n}, \gamma\right) = \frac{\exp\left(\varphi\left[C_{ni}, L_i \middle| \boldsymbol{x_{ni}}, \boldsymbol{\beta_n}, \gamma_i\right]\right)}{\sum_{s \in J_n} \exp\left(\varphi\left[C_{ns}, L_s \middle| \boldsymbol{x_{ns}}, \boldsymbol{\beta_n}, \gamma_s\right]\right)} \tag{4.3}$$

Assuming that individuals take labor market conditions as given, we can rewrite:

$$\mathrm{P}\left(U_{ni} > U_{nj}, \forall j \neq i \middle| \boldsymbol{x_n}, \boldsymbol{\beta_n}, \gamma\right) = \frac{\exp\left(v\left[C_{ni}, L_i \middle| \boldsymbol{x_{ni}}, \boldsymbol{\beta_n}\right]\right) g\left(i \middle| \boldsymbol{x_{ni}}, \gamma_i\right)}{\sum_{s \in J_n} \exp\left(v\left[C_{ns}, L_s \middle| \boldsymbol{x_{ns}}, \boldsymbol{\beta_n}\right]\right) g\left(s \middle| \boldsymbol{x_{ns}}, \gamma_s\right)} \tag{4.4}$$

with $v(C_{nj}, L_j)$ as systematic utility function and $g(j)$ as frequency of feasible jobs with type j. Hence, the individual choice probability is given as the systematic utility part weighted by the availability of jobs with type j. In the following, we discuss the specification of $v(\cdot)$ and $g(\cdot)$ as well as the estimation procedure.

4.2.2 Identification

Econometrically, the discrete choice approach boils down to the representation of the labor supply decision in a random utility model. In the very basic model, the theoretical set-up implies that the household's decision satisfies the Independence of Irrelevant Alternatives (IIA) property (Luce, 1959). In other words, the preference between two alternatives does not depend on the presence of a third one. While this assumption may seem rather restrictive at first glance, Dagsvik and Strøm (2004) and Train (2009) show that it is well in line with economic intuition and even less restrictive than the necessary assumptions to estimate continuous hours models. However, the IIA assumption is no longer needed as soon as additional random effects are incorporated in the model (see Section 4.2.3).

It is crucial to impose a specific functional form for both $v(C_{nj}, L_{nj})$ and $g(j)$ to obtain consistent estimates of $\boldsymbol{\beta_n}$ and $\boldsymbol{\gamma_j}$. van Soest et al. (2002) show that semi-parametric specifications also yield consistent results. As consumption is a function of working hours and thus leisure, identification of preference parameters relies on (i) the variation in working hours h_j, hourly wages w_{nj}, non-labor income I_n, and other characteristics $\boldsymbol{x_{nj}}$, and (ii) the fact that the tax function $f(w_{nj}h_j, I_n)$ is highly non-linear in h_j and w_{nj}. This also implies that labor market conditions $\boldsymbol{\gamma_j}$ can only be separated and identified on the assumption of a specific functional form (Dagsvik and Strøm, 2006).

In addition to this, the vast majority of the literature also assumes that preferences $\boldsymbol{\beta_n}$ and labor market conditions $\boldsymbol{\gamma_j}$ may depend on individual characteristics, but are independent of the wage rate w_{nj}. Thus, it is commonly assumed that:

$$\mathrm{E}\left[\beta_n w_{nj} \middle| \boldsymbol{x}_{nj}\right] = 0 \qquad \mathrm{E}\left[\gamma_j w_{nj} \middle| \boldsymbol{x}_{nj}\right] = 0 \tag{4.5}$$

The main reason for this assumption is that it reduces the computational burden substantially and makes the estimation more convenient.

In order to estimate the preference coefficients, one has to evaluate both functions $v(\cdot)$ and $g(\cdot)$ for every household $n = 1, \ldots, N$ and every choice category within the choice set J_n. Given the different income levels, the model can be estimated via maximum likelihood. The derivation of the (log)-likelihood function is straightforward (McFadden, 1974). However, some modeling assumptions have to be made, as well as several possible extensions to this simple set-up.

4.2.3 Modeling Decisions

Choice Set. The first modeling decision relates to the construction of the choice set. Most authors simply pick a set of representative levels of hours of work and assume (small) identical choice sets for the whole population. In our analysis, we follow this literature and assume that households with a single decision maker face seven possible labor supply states, i.e., either non-participation or working 10, 20, 30, 40, 50 or 60 hours per week. Couple households are assumed to face 7^2 alternatives. The results are generally not sensitive to the number of choices (e.g., 4 vs. 7 vs. 13) or the exact value assigned to each category (see, e.g., Bargain et al., 2014). As noted before, we focus on other aspects of the model set up, namely the specification of the utility function and the treatment of wages. See Aaberge et al. (2009) for a detailed discussion of alternative representations of the choice set.

Functional Form of the Systematic Utility. As the discrete choice approach relies on the comparison of different utility levels, it is crucial to determine the form of the systematic utility function. In theoretical terms, the function $v(\cdot)$ represents the direct utility function of the household. Most applications rely on either a translog, a quadratic or a Box-Cox transformed utility specification. However, several other choices are possible.

Heterogeneity in Preferences. Heterogeneity in the labor supply behavior along observable characteristics can be rather easily introduced in the context of structural labor supply models by extending the utility specification. The preference coefficients of the direct utility function are usually interacted with some observed household characteristics, such as age or the presence of children, as taste shifters.

Additionally accounting for unobserved heterogeneity overcomes the IIA assumption as it allows for unobservable variation in preferences between choice alternatives. There are two main ways to do so: in most applied works, either a *random coefficient model* (van Soest, 1995) or a *latent class model* (Hoynes, 1996) is assumed. The former typically assumes a set of preference coefficients to be (multivariate) normally distributed, whereas the latter allows a set of discrete mass points for the estimated coefficients. Keane and Wasi (2012) discuss the performance of both approaches. We focus on the random coefficient approach as it has become standard in the literature.

Welfare Stigma and Benefit Take-Up. While the model as described thus far assumes that households only build their preferences with respect to the levels of consumption and leisure, their utility may also depend on the *source* of income. For example, the participation in welfare programs may be connected to an unobservable stigma that affects the household's utility and prevents some households from taking up benefits (Moffitt, 1983). In the discrete choice context, this can be incorporated by accounting for the potential disutility from welfare participation and expanding the choice set such that the household explicitly chooses between benefit take-up and non-participation (Hoynes, 1996, Keane and Moffitt, 1998).

Fixed Costs and Hours Restrictions. Moreover, van Soest (1995) argues that working part-time could also be connected with an unobservable disutility, because part-time jobs may exhibit higher search costs. Euwals and van Soest (1999) extend this idea by introducing fixed costs of work, which have since been used in several applications. While both approaches help to explain the observed labor market outcomes, their rationale remains rather *ad hoc.* Aaberge et al. (1995) provide a micro foundation that allows a structural interpretation of fixed costs and the utility connected to certain hours alternatives. In their model, households choose between (latent) job offers that differ not only regarding the working hours, but also in terms of availability, wages, and non-monetary attributes.

4.2.4 Wage Imputation Procedure

In addition to the specification of the utility function, there are important modeling assumptions regarding the wage imputation. In order to calculate the disposable income for the different choice alternatives, one needs information on the hourly wage rates. While for actual workers the wage rate can be calculated by gross earnings and hours of work (we use standardized working hours to reduce the potential division bias, see Borjas, 1980, and Ziliak and Kniesner, 1999, for a discussion), the wage information is typically missing for non-workers. The first decision is how to deal with missing wages in the estimation process. In practice, wages are either estimated beforehand and treated as given within the estimation of the labor supply model or wages and preferences are estimated jointly. In addition, one has to decide whether the estimated wage rates are used only if wages are not observed or for the full sample (see MaCurdy et al., 1990, for a discussion of the advantages and disadvantages of both approaches). In either case, one can ignore or explicitly account for potential sample selection issues in observed wages.

After fitting the wage equation, another important question is whether the potential errors in the wage rate prediction are incorporated in the labor supply estimation. Especially when using predicted wages for the full sample, the "new" distribution of wages will typically have a significantly lower variance and the predicted wage will differ considerably from the observed one, at least for some workers. Thus, ignoring the error when predicting wage rates, which is still done in practice, leads to inconsistent estimates. The standard procedure to incorporate wage prediction errors is to integrate over the estimated wage distribution and thus integrating out the wage prediction error during the estimation (van Soest, 1995). One approximation used in some applications is to simply add a single random draw to the predicted wage rates (Bargain et al., 2014). While this procedure lacks a theoretical rationale, it substantially reduces the computational burden of the estimation.

4.2.5 Estimation Approach

The named extensions – especially regarding the inclusion of unobserved heterogeneity and the incorporation of wage prediction errors – complicate the estimation procedure and lead to the more general representation as *mixed logit model* (Train, 2009). Taking the most general specification as reference, the likelihood function can be written as:

$$L = \prod_{n=1}^{N} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\exp\left(v_{ni}\left[\cdot|\hat{w}_{ni}, \boldsymbol{\beta_n}\right]\right) g\left(i|\gamma_i\right)}{\sum_{j \in J_n} \exp\left(v_{nj}\left[\cdot|\hat{w}_{nj}, \boldsymbol{\beta_n}\right]\right) g\left(j|\gamma_j\right)} f(\boldsymbol{\beta_n}, \gamma) f(\hat{\boldsymbol{w}}_{\boldsymbol{n}})\, \mathrm{d}\boldsymbol{\beta_n} \mathrm{d}\gamma \mathrm{d}\hat{\boldsymbol{w}}_{\boldsymbol{n}} \tag{4.6}$$

where $i \in J_n$ denotes the alternative chosen by individual n. The likelihood contributions not only depend on the systematic utility function, but also on the availability of the choice alternatives, denoted by $g(i)$. This set-up implies that the availability of choice alternatives can be separated from the systematic utility, which is a reasonable assumption at least for labor markets in industrialized countries. As the preferences may also include unobserved heterogeneity, the probability that household n maximizes its utility at choice alternative i has to be integrated over the distribution of coefficients $(\boldsymbol{\beta_n}, \gamma)$. Similarly, the individual likelihood contributions have to be integrated over the range of possible wage predictions $\hat{w}_{nj}$. As both variables will typically not be uniformly distributed, the choice probability has to be weighted by the probability density of the random components.

The model as shown in equation (4.6) is very general and less restrictive than the conditional logit set-up. In turn, it is no longer possible to find an analytical solution. Train (2009) proposes the use of maximum simulated likelihood methods instead. In order to retrieve the simulated likelihood, the double integral has to be approximated and averaged over $r = 1, \ldots, R$ random draws from the distributions of $(\boldsymbol{\beta_n}, \gamma)$ and $\hat{w}_{nj}$. The simulated log-likelihood is given by:

$$\ln(SL) = \sum_{n=1}^{N} \ln\left(\frac{1}{R}\sum_{r=1}^{R} \frac{\exp\left(v_{ni}\left[\cdot\middle|\hat{w}_{ni}^{(r)}, \boldsymbol{\beta}_{\boldsymbol{n}}^{(\boldsymbol{r})}\right]\right) g\left(i\middle|\gamma_i^{(r)}\right)}{\sum_{j\in J_n} \exp\left(v_{nj}\left[\cdot\middle|\hat{w}_{nj}^{(r)}, \boldsymbol{\beta}_{\boldsymbol{n}}^{(\boldsymbol{r})}\right]\right) g\left(j\middle|\gamma_j^{(r)}\right)}\right) \tag{4.7}$$

When the number of draws goes to infinity, the simulated log-likelihood in (4.7) converges to the log-likelihood of the model denoted in (4.6). Instead of relying on conventional random draws, we approximate the likelihood function using pseudo-random Halton sequences. This reduces the number of draws needed to ensure stable results as Halton sequences cover the desired distribution more evenly (Train, 2009).[4]

4.2.6 Common Specifications in the Literature

Tables 4.1 and 4.2 provide an overview on the empirical specification of several popular models that have been applied in recent years and that are used as key references in the literature. Mainly three utility functions have been used, i.e., either a translog, a quadratic or a Box-Cox transformed specification. As the Stone-Geary function can be interpreted as a simplification of the translog or the Box-Cox utility function, only the higher-degree polynomials used in van Soest et al. (2002) stand out from the list. Their approach can be seen as approximation to a non-parametric specification of the utility function. The inclusion of observed heterogeneity shows a similar picture. All studies allow for observed heterogeneity in the preferences for leisure, whereas fewer studies allow for preference heterogeneity regarding consumption. The evidence on unobserved heterogeneity is somewhat more mixed, just like the inclusion of heterogeneity in fixed costs and the potential stigma from welfare participation.

As working hours are typically concentrated in only few hours categories, most authors include fixed costs of working, hours restrictions, or both in their models. Fixed costs and hours restrictions can also be loosely interpreted as measures for the availability of the respective choice alternatives (Aaberge et al., 2009). Less than half of the models explicitly allow for stigma effects and non-take-up of welfare benefits. This is interesting due to the common finding that the actual benefit participation rate deviates substantially from full take-up. Thus, models that do not account for the potential disutility are expected to over-predict the number of recipients.

[4] Details on the estimation procedure can be found in Löffler (2013).

Table 4.1: Model Specifications

Paper	Utility Function	Heterogeneity* Observed	Heterogeneity* Unobs.	Welfare Stigma	Constraints
Aaberge et al. (1995, 2009)	Box-Cox	L	–	–	FC, HR
Aaberge et al. (1999)	Box-Cox	L, FC	–	–	FC, HR
Dagsvik and Strøm (2006)	Box-Cox	L, FC	–	–	FC, HR
Dagsvik et al. (2011)	Box-Cox	L, FC	–	–	FC, HR
Blundell and Shephard (2012)	Box-Cox	L, C, S, FC	C, S	Yes	FC
van Soest (1995)	Translog	L	$-/L$†	–	$-/HR$
Euwals and van Soest (1999)	Translog	L, FC	L	–	FC
van Soest and Das (2001)	Translog	L, FC	L	–	FC
Flood et al. (2004)	Translog	L, L^2, S	L, L^2, S	Yes	–
Haan (2006)	Translog	L, C	$-/C$	–	HR
Flood et al. (2007)	Translog	L, C, FC, S	L, C, FC, S	Yes	FC
Hoynes (1996)	Stone-Geary	L, S	L, S	Yes	$-/FC$
van Soest et al. (2002)	Polynomial	L	L	–	FC
Keane and Moffitt (1998)	Quadratic	L, S	L, S	Yes	–
Blundell et al. (1999, 2000)	Quadratic	L, C, FC	C, S	Yes	FC
Bargain et al. (2014)	Quadratic	L, C, FC	C	–	FC

* L and C denote heterogeneity in preferences for leisure and consumption, respectively. S denotes the disutility from welfare participation. FC refers to fixed costs of working and HR to hours restrictions.
† Robustness checks and alternative model specifications are separated by slashes.

Table 4.2: Wage Imputation Methods

Paper	Estimation Approach	Sample Selection	Imputation	Prediction Error
Aaberge et al. (1995, 2009)	Simultaneous	–	Full sample	–
Aaberge et al. (1999)	Simultaneous	–	Full sample	–
Keane and Moffitt (1998)	Simult./Two step*	–	Non-workers	–
van Soest et al. (2002)	Simultaneous	–	Non-workers	Integrated out
Blundell and Shephard (2012)	Simult./Two step	–	Non-workers	Integrated out
van Soest (1995)	Two step	Yes	Non-workers	–/Integrated out
Euwals and van Soest (1999)	Two step	Yes	Non-workers	Integrated out
Blundell et al. (1999, 2000)	Two step	Yes	Non-workers	Integrated out
van Soest and Das (2001)	Two step	Yes	Non-workers	Integrated out
Haan (2006)	Two step	Yes	Non-workers	–
Flood et al. (2007)	Two step	Yes	Non-workers	–/Integrated out
Dagsvik et al. (2011)	Two step	Yes	Non-workers	–
Hoynes (1996)	Two step	Yes	Full sample	–
Flood et al. (2004)	Two step	Yes	Full sample	–
Dagsvik and Strøm (2006)	Two step	Yes	Full sample	Integrated out
Bargain et al. (2014)	Two step	Yes	Full sample	Random draw

* Robustness checks and alternative model specifications are separated by slashes.

Less variation can be found in terms of the model's treatment of wages. While most studies estimate wages and the labor supply decision separately in a two-step procedure, only the models of Aaberge et al. (1995, and follow-ups), Keane and Moffitt (1998), van Soest et al. (2002) and Blundell and Shephard (2012) apply a simultaneous maximum likelihood estimation. In turn, these models neglect potential sample selection issues when estimating wages. There is no consensus in the literature whether predicted wages should be used only for individuals whose wages are unobserved or for the full sample. Regarding the handling of the wage prediction errors, it becomes increasingly common practice to incorporate and integrate out the errors during the estimation.

4.3 Data

The baseline estimations in this paper are performed on the German Socio-Economic Panel (SOEP), a representative household panel survey for Germany (Wagner et al., 2007). SOEP includes now more than 24,000 individuals in around 11,000 households. We use the 2008 wave of the SOEP, which provides household data from 2008, as well as data on the labor supply behavior and incomes from the preceding year (i.e., the year before the Great Recession). We rely on the tax and transfer system of 2007, focusing our analysis on the working age population and thus excluding individuals younger than 17 or above the retirement age of 65 from our estimations. Our sample is further restricted to those households where at least one decision maker can freely adjust her labor supply. Therefore, we exclude households where all decision makers are self-employed (since it is difficult to measure true hours and wages for those), civil servants[5] or in the military service. Moreover, our subsample includes some households with more than two adults, which mainly includes adult children living with their parents. We exclude these young adults from the estimation as it is unclear how their consumption and utility are determined (Dagsvik et al., 2011).

As labor supply is known to be rather heterogeneous across population subgroups, we split the sample into five distinct demographic subpopulations ("labor supply types"). The first two groups are defined as single men and single women with or without dependent children. Our estimation sample contains 779 households with single males and 1,065 households with single females. In addition, we specify three different kinds of couple households. First, we define 688 couple households in which the male partner has a flexible labor supply but the female partner is inflexible (e.g., due to self-employment or exclusion restrictions regarding the age). Second, we have 1,042 couple households in which the male partner has an inflexible labor

[5] Tenured civil servants cannot freely adjust the weekly working hours. Note that we keep all other public sector employees.

supply but the female partner is flexible. In order to model the household labor supply decision of these "semi-flexible" couple households, we assume that the flexible partner faces his or her labor supply decision conditional on the labor supply behavior of the inflexible partner. Third, our sample includes 3,099 couple households in which both partners are flexible with respect to their labor supply behavior.

For the computation of consumption levels for the different choice categories, we rely on IZA's policy simulation model IZAΨMOD (v3.0.0), which incorporates a very detailed representation of the German tax and benefit system (see Löffler et al., 2014a, for a comprehensive documentation). Some of the estimated models would require applying the tax and benefit system for every possible wage rate for every household in every step of the numerical likelihood maximization. To avoid this cumbersome procedure, we approximate the tax and benefit system by using a highly flexible second-degree polynomial that transforms monthly gross earnings into disposable income while controlling for a rich set of household characteristics, as well as all available sources of non-labor income. The resulting R^2 shows a very good fit of more than 99 percent for all population subgroups but single women (only 97 percent for them), which confirms that our approximation performs rather well.[6] The results are very much in line with those taking advantage of the full representation of the tax and transfer system, we are thus confident that the approximation does not affect our findings.

As a robustness check, we compare our results obtained with German data to results for the US. For this, we use data from IPUMS-CPS which is an integrated data set of the March Current Population Survey (CPS) for 2007. In order to calculate income and payroll taxes, we use NBER's simulation model TAXSIM.

4.4 Meta-Analysis of Labor Supply Models

Robustness checks in the applied labor supply literature usually narrow down to a small deviation in just one of the modeling assumptions (see Tables 4.1 and 4.2). Evers et al. (2008) and Bargain and Peichl (2016) perform meta-analyses of labor supply models comparing estimated labor supply elasticities for different countries and explain them mainly by study characteristics. In either case, it is difficult to draw general conclusions on the exact specification of discrete choice models from the reported results. We overcome these difficulties by estimating a large variety of different modeling assumptions in a controlled environment using the same data. The estimation results allow us to determine how sensitive the estimated outcomes are with respect to the specification and the wage imputation procedure used in the model.

[6] We combine the predicted amounts of consumption with a single random draw for each household; otherwise, we would mistakenly reduce the variance in the consumption variable.

4.4.1 Set Up of the Analysis

For our analysis, we combine frequently used modeling assumptions and estimate all sensible combinations of these specifications. We estimate 3,456 different model specifications for the five distinct population groups (see Section 4.3), which leads us to 17,280 maximum likelihood estimations. However, the sample of estimation results is reduced because not all models did converge to a global maximum in a reasonable time span. We drop those estimation results from our analysis that did not converge after 100 iterations of Stata's maximum likelihood implementation. Depending on the labor supply group we lose up to six percent of our sample and end up with 16,730 different estimation results.[7]

Table 4.3 shows the different specifications and the number of converged estimation results. The table reads as follows. We estimate 1,152 distinct models with a Box-Cox transformed utility specification for each of the five labor supply groups. Because few models did not converge to a global maximum in a reasonable amount of time, only 1,022 estimation results for single males and 1,132 for single females are included in our sample. Regardless of the functional form of the utility function, 1,152 of the estimated models neglect any kind of hours restrictions or fixed costs, 1,152 models include part-time restrictions and 1,152 models account for fixed costs of work.

To make the estimation results comparable across the different labor supply groups, we standardize the statistical fit and the estimated elasticities within population groups. We subsequently pool the data and regress the estimation results on indicators for the different modeling assumptions (mainly represented as dummy variables). We measure the statistical fit by the Akaike Information Criterion (AIC) of the models. To retrieve (uncompensated) labor supply elasticities, we increase the own-wage rates by ten percent and simulate the labor supply reaction to this wage change.[8]

4.4.2 Empirical Results

The results of these meta-regressions can be found in Table 4.4. Coefficients have to be compared to the simple reference model using a translog utility function, neglecting observed and unobserved heterogeneity in preferences as well as fixed costs of working, hours restrictions or any stigma from welfare participation. In this reference model, we use observed wage rates for actual workers and predict wages for non-workers without incorporating the wage prediction error in the labor supply estimation. All outcomes are standardized, i.e., coefficients relate to changes in terms of standard deviations, and thus only large estimates (in absolute values) are

[7] Of course, more complex models take longer to converge. Apart from that, we do not find systematic effects of different types of assumptions on the probability to converge.

[8] Results are robust to different ways of computing own-wage labor supply elasticities, see below for details.

Table 4.3: Estimated Model Combinations

			Number of Converged Models				
			Singles		Couples		
Model Parameter	Option	All (1)	Male (2)	Female (3)	Male (4)	Female (5)	Both (6)
Utility Function	Box-Cox	1,152	1,022	1,132	951	1,148	1,029
	Quadratic	1,152	1,152	1,151	1,152	1,133	1,152
	Translog	1,152	1,125	1,144	1,148	1,148	1,143
Welfare Stigma	No	1,728	1,642	1,701	1,607	1,713	1,664
	Yes	1,728	1,657	1,726	1,644	1,716	1,660
Hours Restrictions	–	1,152	1,091	1,141	1,040	1,131	1,109
	Fixed Costs	1,152	1,064	1,137	1,061	1,149	1,063
	Part-Time	1,152	1,144	1,149	1,150	1,149	1,152
Number of Halton Draws	–	288	288	288	283	288	286
	10 Draws	1,584	1,440	1,564	1,429	1,559	1,456
	5 Draws	1,584	1,571	1,575	1,539	1,582	1,582
Observed Heterogeneity	–	864	835	864	822	860	834
	In β_C Only	864	827	862	834	861	822
	In β_L Only	864	827	858	798	859	836
	In β_L, C	864	810	843	797	849	832
Unobserved Heterogeneity	–	576	574	571	566	570	574
	In β_C Only	864	863	853	846	862	863
	In β_L Only	576	520	574	523	569	541
	In β_L, β_C	864	804	856	795	854	791
	With Correl.	576	538	573	521	574	555
Wage Imputation	Full Sample	1,728	1,652	1,708	1,635	1,710	1,655
	Non-Workers	1,728	1,647	1,719	1,616	1,719	1,669
Wage Prediction Error	–	1,296	1,217	1,293	1,219	1,291	1,245
	1 Random Draw	1,296	1,236	1,291	1,203	1,284	1,239
	Integrated Out	864	846	843	829	854	840
Total		3,456	3,299	3,427	3,251	3,429	3,324

Notes: This table shows the number of estimated models over the different model parameters and population subgroups. Column (1) shows the number of possible model combinations for each choice of parameters. Columns (2)-(6) report the number of converged models by population subgroup. Column (2) refers to single male households, column (3) to households with a single female adult (both also including lone parents). Columns (4)-(6) refer to couple households where only the male partner is flexible in his labor supply behavior, where only the female partner is flexible, or where both partners are flexible in their labor supply, respectively.

Table 4.4: Marginal Impact of Modeling Assumptions (SOEP)

	Fit	10 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Quadratic	0.119***	0.124***	-0.015	0.004
	(0.023)	(0.028)	(0.062)	(0.053)
Box-Cox	-0.020	0.116***	0.080**	0.085**
	(0.026)	(0.040)	(0.035)	(0.034)
Welfare Stigma	0.968***	0.045	0.065	0.065
	(0.076)	(0.062)	(0.047)	(0.042)
Number of Halton Draws	-0.010***	0.005	-0.003	-0.002
	(0.001)	(0.004)	(0.004)	(0.004)
Hours Restrictions				
Part-Time Restrictions	-1.647***	0.384***	0.105**	0.152***
	(0.082)	(0.070)	(0.039)	(0.042)
Fixed Costs	-1.093***	0.481***	0.187***	0.238***
	(0.070)	(0.067)	(0.040)	(0.041)
Observed Heterogeneity				
In β_C Only	-0.335***	-0.049	0.060**	0.043*
	(0.057)	(0.035)	(0.022)	(0.023)
In β_L Only	-0.381***	0.048	0.045**	0.046*
	(0.061)	(0.038)	(0.021)	(0.023)
In β_C and β_L	-0.475***	0.016	0.012	0.013
	(0.070)	(0.044)	(0.019)	(0.022)
Unobserved Heterogeneity				
In β_C Only	0.005	-0.006	-0.059*	-0.051
	(0.014)	(0.023)	(0.032)	(0.030)
In β_L Only	0.005	-0.081***	-0.029	-0.037
	(0.013)	(0.023)	(0.028)	(0.027)
In β_C And β_L	-0.041***	-0.037	-0.069**	-0.064**
	(0.013)	(0.024)	(0.027)	(0.026)
In β_C And β_L With Correlation	-0.119***	-0.082**	-0.102***	-0.101***
	(0.016)	(0.034)	(0.034)	(0.033)
Wage Imputation				
Full Sample, No Correction	-0.811***	2.121***	2.235***	2.240***
	(0.119)	(0.094)	(0.091)	(0.086)
Full Sample, Error Integrated Out	-0.530***	1.399***	1.385***	1.406***
	(0.048)	(0.119)	(0.123)	(0.124)
Full Sample, 1 Random Draw	-0.104**	0.071	0.131	0.121
	(0.049)	(0.062)	(0.093)	(0.088)
Non-Workers, Error Integrated Out	0.000	0.048	0.040	0.041
	(0.067)	(0.063)	(0.041)	(0.041)
Non-Workers, 1 Random Draw	0.070	-0.230***	-0.232***	-0.235***
	(0.056)	(0.038)	(0.035)	(0.037)
Constant	1.004***	-0.939***	-0.678***	-0.726***
	(0.121)	(0.094)	(0.087)	(0.087)
Labor Supply Type Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,730	13,219	13,219	13,219
Explanatory Power R^2	0.854	0.849	0.870	0.881

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 10 percent and aggregating individual labor supply responses. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model – the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

also economically interesting. Our results show, e.g., that combining this model set-up with a quadratic utility function instead of a translog specification increases the AIC by 12 percent of a standard deviation in the sample and thereby worsens the statistical fit. We summarize the key findings below.[9]

Goodness of Fit. Although the statistical fit is usually not the main outcome of interest, our results show several interesting patterns for future applications (see Table 4.4). First, the choice of the utility function does not substantially improve or worsen the statistical fit. Our analysis confirms the usual finding that the implementation of hours restrictions, fixed costs and observed preference heterogeneity clearly help to explain the observed labor supply choices, i.e., the AIC decreases. Estimating random coefficients models that also allow for unobserved heterogeneity yields little value-added in terms of the statistical fit – especially compared to the increased computational burden of the estimation. The results regarding the wage imputation show that these modeling decisions affect the statistical fit of the model substantially. Predicting wages not only for non-workers but for the full sample of workers improves the fit significantly. This mirrors the fact that much of the variation in the data is lost by using predicted instead of actual wages for the full sample when not accounting for errors in the wage rate prediction.

More generally, our results show that apart from the implementation of fixed costs or hours restrictions, there is hardly a single modeling assumption that guarantees a good fit. Instead, several small issues help to explain the observed labor market outcomes and add up to a good fit.

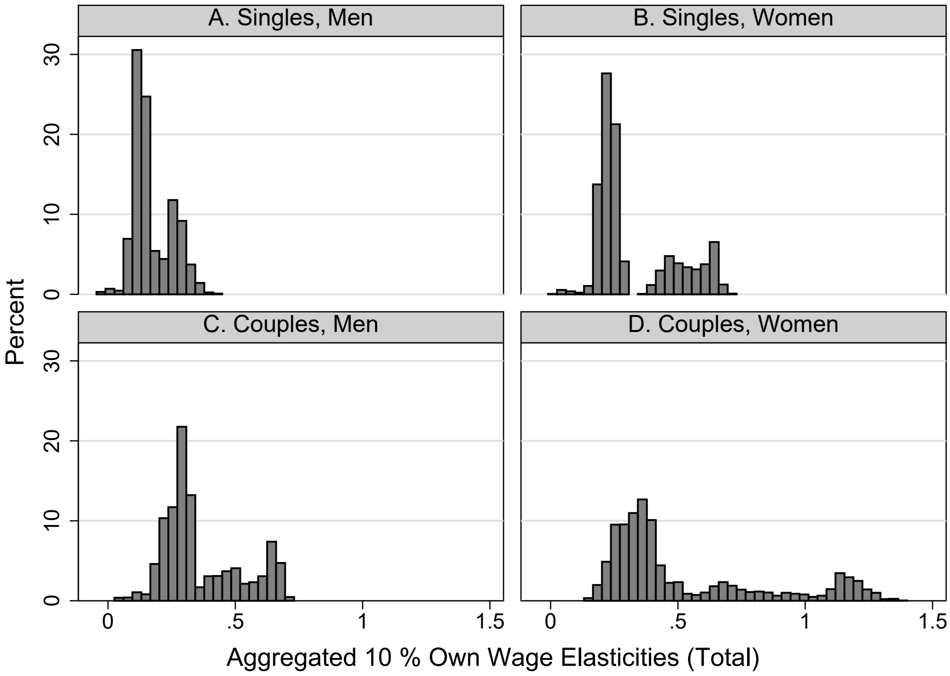
Labor Supply Elasticities. More important than the statistical fit is whether specific modeling assumptions systematically influence the out-of-sample predictions when simulating policy or wage changes. Figure 4.1 shows the distribution of simulated labor supply elasticities across the converged models for four demographic subgroups.[10] The graph shows considerable variation across the different modeling set-ups (within population groups as well as across groups).

In line with the literature, we find that the simulated elasticities are rather robust regarding the specification of the utility function, as well as the implementation of observed and unobserved heterogeneity. This is reassuring as it shows that the frequently applied specifications do not restrict the labor supply decision *a priori*. The only (weak) exception seems to be the

[9] The presented standard errors do not account for the (potential) variation in the statistical fit and the simulated elasticities for one specific model when estimating the same model using different samples. Accounting for this uncertainty, e.g., by using bootstrap procedures, would produce larger standard errors than those presented but is computationally infeasible in our context due to the large number of estimated models.

[10] We aggregated couples with one and couples with two flexible partners in this figure.

Figure 4.1: Simulated Labor Supply Elasticities For Converged Models



Notes: This figure shows the distribution of estimated labor supply elasticities over the 3,456 different model specifications for four labor supply groups (see panels). Elasticities are defined as hours responses to a ten percent increase in the own wage rate, combining both intensive and extensive margin and aggregating over individual responses. Panel A shows elasticities for single men, Panel B shows the results for single women. We pool estimation results for the three types of couple households and plot the response of the male and female partner in Panel C and Panel D, respectively (see Section 4.3 for a discussion of the different labor supply types).

implementation of hours restrictions or fixed costs, which tend to drive extensive elasticities up. This finding supports the view that jobs with very few weekly working hours are harder to find than regular part-time or full-time jobs. It is thus more likely that people switch from non-participation to 20 or 40 than to 5 or 10 hours of work when accounting for this fact, which leads to higher elasticities at the extensive margin.

Substantially more of the variation in simulated elasticities can be explained when analyzing the impact of the wage imputation and the handling of wage prediction errors. Our results thus hold the important message that this part of the model specification is much more relevant to the estimated elasticities than the utility specification. For instance, using predicted wages not only for non-workers but for the full sample of individuals roughly doubles the estimated elasticities. The average own-wage elasticity in our meta-analysis increases from 0.23 to 0.46. This substantial difference can be explained by the fact that predicting wages for the full sample reduces the variance of the wage distribution substantially. To explain the observed working hours with less variation in wages and thus income and consumption, the implied elasticities have to increase. Accounting for wage prediction errors and integrating them out during the estimation markedly reduces the difference. Predicting wages for all individuals but ignoring the wage prediction error yields an average elasticity of 0.65 in our meta-analysis sample as opposed to 0.35 when accounting for the prediction error. Interestingly, the results differ substantially depending on whether a single random draw or higher numbers are used. The *ad hoc* procedure of adding a single random draw tends to cancel out the effect of a full sample prediction, estimated elasticities are close to those of the reference model relying mostly on observed wages (average elasticity of 0.26). In contrast, correcting for the wage prediction error tends to reduce the elasticities, but we still observe the estimated elasticities to be significantly higher than those in which the wage rates were only imputed for non-workers (0.47 vs. 0.25).

Robustness. We perform a wide range of robustness checks to confirm that our results are not driven by the used data or the meta-analysis set up. In particular, we also use a different wave from the same data set and perform our analysis also using data from the CPS for the US (see Table 4.A.1 in the Appendix). In addition to the marginal impact (holding all other specification details constant), we investigate the partial impact of the modeling assumptions (see Table 4.A.2 in the Appendix), which only shows the differences in means due to the specific assumptions (e.g., the mean of elasticities using a translog utility specification vs. the mean of elasticities using different functional forms, irrespective of other modeling issues). The results we obtain are qualitatively the same. We also check the robustness regarding the calculation of elasticities and find no differences whether we simulate one percent or ten percent changes in the own-wage rate (see Table 4.A.3 in the Appendix). Also switching the calculation of the elasticities from aggregated to mean, median or other quantile measures did not affect our findings (see Tables 4.A.4 and 4.A.5 in the Appendix).

Summary. Our results partly confirm previous findings in the literature. While the empirical specification of the systematic utility function has an impact on the statistical fit, we find only little differences in the estimated elasticities. It may thus be justified to rely on simpler model set-ups when the computational burden is a major concern. However, the majority of the robustness checks applied in the literature focus on the effects of different utility specifications and usually ignore how the underlying wage distribution – and especially the imputation of wages – may influence the results. We find that these assumptions explain much more of the variation in simulated labor supply elasticities than the specification of the utility function. Most previous robustness checks have thus concentrated on issues of secondary order. Instead, more attention should be paid to the wage imputation and the handling of wage prediction errors. Modeling choices regarding the wage handling may thus also explain part of the large variation found in labor supply studies.

Which assumption should be preferred? Integrating out the error term of the wage prediction is clearly preferred over no correction. Thanks to advances in computing power, the additional computational burden should not be an issue anymore. Using only one random draw from the wage distribution, which has been used as a shortcut to avoid long computations, is hence not necessary anymore. In terms of predicting wages for the full sample vs. non-workers only, the answer depends on the research question and data at hand. The first option assumes that all individuals, not only the unemployed, are aware of uncertainties about their individual wage realization and base their labor supply decision on expected wages as derived from the Mincerian wage equation. The second option, on the other hand, assumes that employed workers make choices based on their current wage rate, independent of whether they drew a positive, negative or no wage shock in their current job. Which of these models fits better is a decision that the researcher has to make and it should be made explicit.

4.5 Conclusion

Structural labor supply models are frequently used in the empirical labor supply analysis for many different purposes. In recent years, it has become a standard procedure to estimate labor supply decisions as a choice among a set of different hours alternatives or job opportunities. In contrast to this popularity, little is known about how the numerous modeling assumptions influence the statistical fit as well as the simulated labor supply elasticities.

In this paper, we provide an overview of the most important specification issues and conduct a comprehensive sensitivity analysis to disentangle the driving factors behind the results obtained from structural labor supply models. Our results show that even if the modeling assumptions concerning the direct utility specification increase or worsen the statistical fit, i.e., the power to explain the observed labor supply behavior, the models are robust in terms of their implied labor supply elasticities. In contrast, the model predictions are highly sensitive to changes in the underlying wage distribution, a mechanism almost completely neglected in the literature to date. Thus, the questions of whether to use predicted or observed wages for actual workers and whether and how to integrate the wage prediction error out during the estimation process have a large and statistically significant impact on the statistical fit of the model *and* the estimated labor supply elasticities.

Our findings have important implications for tax policy design. Diamond and Saez (2011) derive simple formulas for the optimal (top) marginal tax rates based upon labor supply elasticities.[11] They assume an elasticity of 0.25 as an "a mid-range estimate from the empirical literature" which is close to our mean estimate for models using the observed wage distribution. This leads to an optimal top marginal tax rate of $\tau = \frac{1}{1+1.5 \cdot 0.25} = 72.7$ percent. However, an elasticity of 0.65 as found in models using predicted wages reduces the optimal tax rate to 50.6 percent bringing it closer to actually observed values (the top labor tax rate in the US is 42.5 percent). While we cannot claim that we have identified the true value for the labor supply elasticity – which might not even exist – our analysis shows that more attention should be paid to the specification of structural labor supply models when using them for policy analysis. Future research should try estimating preferences and wages jointly.

[11] The formula for the optimal top tax rate is $\tau = \frac{1-g}{1-g+a \cdot e}$ where g is the marginal social welfare weight for the top earners, a is the parameter of the Pareto (income) distribution and e is the labor supply elasticity. Diamond and Saez (2011) assume $g = 0$ to derive the optimal revenue maximizing top tax rate and use an estimated Pareto coefficient of $a = 1.5$ for the US.

Appendix 4.A Additional Results

Table 4.A.1: Marginal Impact of Modeling Assumptions (CPS)

	Fit	10 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Quadratic	0.640***	0.217	0.207	0.210
	(0.062)	(0.183)	(0.185)	(0.185)
Number of Halton Draws	-0.015***	0.022**	0.023**	0.022**
	(0.002)	(0.008)	(0.008)	(0.008)
Hours Restrictions				
Part-Time Restrictions	-1.855***	0.420**	0.397*	0.403*
	(0.089)	(0.188)	(0.189)	(0.190)
Fixed Costs	-1.279***	0.192	0.125	0.142
	(0.067)	(0.120)	(0.104)	(0.106)
Observed Heterogeneity				
In β_C Only	-0.138***	-0.152***	-0.051	-0.078**
	(0.015)	(0.036)	(0.031)	(0.031)
In β_L Only	-0.258***	-0.066*	-0.080*	-0.076*
	(0.026)	(0.036)	(0.044)	(0.042)
In β_C And β_L	-0.309***	-0.115**	-0.094*	-0.097**
	(0.027)	(0.040)	(0.044)	(0.043)
Unobserved heterogeneity				
In β_C Only	0.067***	-0.118***	-0.113**	-0.114**
	(0.013)	(0.037)	(0.044)	(0.043)
In β_L Only	0.070***	-0.119**	-0.121**	-0.120**
	(0.007)	(0.053)	(0.054)	(0.055)
In β_C And β_L	0.046***	-0.089**	-0.083**	-0.084**
	(0.009)	(0.032)	(0.036)	(0.035)
In β_C And β_L With Correlation	0.021***	-0.063**	-0.058**	-0.059**
	(0.007)	(0.029)	(0.025)	(0.026)
Wage Imputation				
Full Sample, No Correction	-0.111*	0.912***	0.918***	0.921***
	(0.057)	(0.284)	(0.300)	(0.299)
Full Sample, 1 Random Draw	0.025	0.338	0.428*	0.413*
	(0.046)	(0.244)	(0.229)	(0.230)
Non-Workers, 1 Random Draw	0.030	-0.329	-0.237	-0.255
	(0.053)	(0.362)	(0.362)	(0.363)
Constant	0.832***	-0.813***	-0.864***	-0.857***
	(0.098)	(0.271)	(0.270)	(0.270)
Labor Supply Types Fixed Effects	Yes	Yes	Yes	Yes
Observations	4,305	3,439	3,439	3,439
Explanatory Power R^2	0.820	0.353	0.340	0.344

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 10 percent and aggregating individual labor supply responses. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model – the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

Table 4.A.2: Partial Impact of Modeling Assumptions (SOEP)

	Fit	10 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Translog	-0.045*	-0.125***	-0.035	-0.047
	(0.024)	(0.021)	(0.045)	(0.040)
Quadratic	0.135***	0.067*	-0.054	-0.037
	(0.013)	(0.039)	(0.053)	(0.046)
Box-Cox	-0.093***	0.061	0.094**	0.090**
	(0.017)	(0.049)	(0.034)	(0.034)
Welfare Stigma	0.965***	0.051	0.072	0.071
	(0.076)	(0.061)	(0.047)	(0.042)
Number of Halton Draws	-0.013***	0.008	-0.003	-0.001
	(0.003)	(0.007)	(0.007)	(0.007)
Hours Restrictions				
None	1.376***	-0.425***	-0.139***	-0.188***
	(0.075)	(0.067)	(0.038)	(0.039)
Part-Time Restrictions	-1.110***	0.145***	0.013	0.035
	(0.052)	(0.041)	(0.024)	(0.026)
Fixed Costs	-0.244***	0.278***	0.127***	0.153***
	(0.034)	(0.033)	(0.024)	(0.023)
Observed Heterogeneity				
None	0.398***	-0.002	-0.035*	-0.030
	(0.063)	(0.038)	(0.019)	(0.021)
In β_C Only	-0.046**	-0.070***	0.042***	0.024*
	(0.017)	(0.015)	(0.014)	(0.013)
In β_L Only	-0.121***	0.067***	0.028*	0.035**
	(0.020)	(0.015)	(0.014)	(0.014)
In β_C And β_L	-0.235***	0.004	-0.036***	-0.030***
	(0.031)	(0.022)	(0.010)	(0.010)
Unobserved Heterogeneity				
None	0.057	0.090	0.125	0.122
	(0.040)	(0.110)	(0.117)	(0.117)
In β_C Only	0.029*	0.075**	0.013	0.023
	(0.015)	(0.036)	(0.038)	(0.038)
In β_L Only	0.050	-0.123	-0.032	-0.047
	(0.040)	(0.110)	(0.110)	(0.111)
In β_C And β_L	-0.035**	0.039	0.006	0.011
	(0.015)	(0.039)	(0.038)	(0.038)
In β_C And β_L With Correlation	-0.102**	-0.128	-0.124	-0.127
	(0.039)	(0.102)	(0.111)	(0.110)
Wage Imputation				
Full Sample Imputation	-0.498***	1.248***	1.313***	1.317***
	(0.100)	(0.288)	(0.294)	(0.296)
Error Integrated Out	-0.037	0.267	0.190	0.207
	(0.125)	(0.351)	(0.359)	(0.362)
Full Sample, No Correction	-0.720***	1.921***	2.033***	2.036***
	(0.119)	(0.145)	(0.144)	(0.142)
Full Sample, Error Integrated Out	-0.334***	1.004***	0.935***	0.960***
	(0.081)	(0.239)	(0.253)	(0.254)
Full Sample, 1 Random Draw	0.143	-0.599**	-0.554**	-0.569**
	(0.089)	(0.237)	(0.258)	(0.257)
Non-Workers, Error Integrated Out	0.269***	-0.544**	-0.606**	-0.602**
	(0.094)	(0.227)	(0.230)	(0.231)
Observations	16,730	13,219	13,219	13,219

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 10 percent and aggregating individual labor supply responses. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model – the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure ($^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$).

Table 4.A.3: Marginal Impact, Aggregated 1 % Elasticities (SOEP)

	Fit	1 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Quadratic	0.119***	0.132***	0.030	0.043
	(0.023)	(0.028)	(0.050)	(0.043)
Box-Cox	-0.020	0.133***	0.087**	0.094**
	(0.026)	(0.042)	(0.035)	(0.035)
Welfare Stigma	0.968***	-0.028	0.084	0.071
	(0.076)	(0.084)	(0.053)	(0.044)
Number of Halton Draws	-0.010***	0.007*	-0.002	-0.001
	(0.001)	(0.004)	(0.004)	(0.004)
Hours Restrictions				
Part-Time Restrictions	-1.647***	0.390***	0.134***	0.176***
	(0.082)	(0.071)	(0.038)	(0.041)
Fixed Costs	-1.093***	0.494***	0.217***	0.264***
	(0.070)	(0.068)	(0.039)	(0.041)
Observed Heterogeneity				
In β_C Only	-0.335***	-0.057	0.060**	0.042*
	(0.057)	(0.035)	(0.022)	(0.022)
In β_L Only	-0.381***	0.032	0.041*	0.041*
	(0.061)	(0.037)	(0.022)	(0.023)
In β_C And β_L	-0.475***	-0.002	0.016	0.013
	(0.070)	(0.044)	(0.020)	(0.022)
Unobserved Heterogeneity				
In β_C Only	0.005	-0.009	-0.054*	-0.048
	(0.014)	(0.024)	(0.031)	(0.030)
In β_L Only	0.005	-0.085***	-0.032	-0.040
	(0.013)	(0.024)	(0.028)	(0.027)
In β_C And β_L	-0.041***	-0.036	-0.068**	-0.064**
	(0.013)	(0.025)	(0.027)	(0.026)
In β_C And β_L With Correlation	-0.119***	-0.087**	-0.092**	-0.093***
	(0.016)	(0.033)	(0.033)	(0.033)
Wage Imputation				
Full Sample, No Correction	-0.811***	2.089***	2.245***	2.248***
	(0.119)	(0.117)	(0.088)	(0.084)
Full Sample, Error Integrated Out	-0.530***	1.427***	1.398***	1.425***
	(0.048)	(0.103)	(0.114)	(0.117)
Full Sample, 1 Random Draw	-0.104**	0.086	0.102	0.100
	(0.049)	(0.085)	(0.078)	(0.079)
Non-Workers, Error Integrated Out	0.000	0.054	0.046	0.048
	(0.067)	(0.062)	(0.035)	(0.038)
Non-Workers, 1 Random Draw	0.070	-0.157***	-0.220***	-0.214***
	(0.056)	(0.056)	(0.028)	(0.032)
Constant	1.004***	-0.930***	-0.730***	-0.770***
	(0.121)	(0.104)	(0.084)	(0.087)
Labor Supply Type Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,730	13,219	13,219	13,219
Explanatory Power R^2	0.854	0.816	0.880	0.889

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 1 percent and aggregating individual labor supply responses. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model – the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

Table 4.A.4: Marginal Impact, Mean 10 % Elasticities (SOEP)

	Fit	10 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Quadratic	0.119***	0.100***	0.022	0.041
	(0.023)	(0.033)	(0.047)	(0.038)
Box-Cox	-0.020	0.101**	0.097***	0.096***
	(0.026)	(0.041)	(0.034)	(0.033)
Welfare Stigma	0.968***	-0.026	0.034	0.026
	(0.076)	(0.063)	(0.045)	(0.039)
Number of Halton Draws	-0.010***	0.005	-0.003	-0.001
	(0.001)	(0.004)	(0.004)	(0.004)
Hours Restrictions				
Part-Time Restrictions	-1.647***	0.353***	0.169***	0.219***
	(0.082)	(0.089)	(0.045)	(0.056)
Fixed Costs	-1.093***	0.448***	0.254***	0.307***
	(0.070)	(0.087)	(0.044)	(0.053)
Observed Heterogeneity				
In β_C Only	-0.335***	0.037	0.048*	0.046
	(0.057)	(0.036)	(0.024)	(0.027)
In β_L Only	-0.381***	0.187***	0.068***	0.101***
	(0.061)	(0.040)	(0.022)	(0.027)
In β_C And β_L	-0.475***	0.187***	0.036	0.074**
	(0.070)	(0.053)	(0.022)	(0.029)
Unobserved Heterogeneity				
In β_C Only	0.005	0.001	-0.046	-0.036
	(0.014)	(0.022)	(0.031)	(0.029)
In β_L Only	0.005	-0.075***	-0.031	-0.041
	(0.013)	(0.026)	(0.028)	(0.026)
In β_C And β_L	-0.041***	-0.027	-0.059**	-0.053**
	(0.013)	(0.025)	(0.026)	(0.025)
In β_C And β_L With Correlation	-0.119***	-0.083**	-0.097***	-0.098***
	(0.016)	(0.036)	(0.033)	(0.033)
Wage Imputation				
Full Sample, No Correction	-0.811***	2.130***	2.264***	2.267***
	(0.119)	(0.106)	(0.092)	(0.089)
Full Sample, Error Integrated Out	-0.530***	1.265***	1.365***	1.364***
	(0.048)	(0.134)	(0.132)	(0.140)
Full Sample, 1 Random Draw	-0.104**	0.058	0.146	0.122
	(0.049)	(0.049)	(0.087)	(0.078)
Non-Workers, Error Integrated Out	0.000	0.035	0.062	0.057
	(0.067)	(0.053)	(0.042)	(0.043)
Non-Workers, 1 Random Draw	0.070	-0.164***	-0.220***	-0.210***
	(0.056)	(0.028)	(0.034)	(0.035)
Constant	1.004***	-0.966***	-0.749***	-0.814***
	(0.121)	(0.119)	(0.091)	(0.097)
Labor Supply Type Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,730	13,219	13,219	13,219
Explanatory Power R^2	0.854	0.820	0.876	0.883

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 10 percent and taking the mean individual labor supply response. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model – the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

Table 4.A.5: Marginal Impact, Median 10 % Elasticities (SOEP)

	Fit	10 % Own Wage Elasticities		
	AIC	Ext.	Int.	Total
Utility Function				
Quadratic	0.119***	0.107***	0.079	0.098*
	(0.023)	(0.030)	(0.063)	(0.048)
Box-Cox	-0.020	0.084**	0.042	0.056
	(0.026)	(0.036)	(0.044)	(0.040)
Welfare Stigma	0.968***	0.001	0.062	0.059
	(0.076)	(0.059)	(0.068)	(0.062)
Number of Halton Draws	-0.010***	0.005	-0.005	-0.003
	(0.001)	(0.005)	(0.005)	(0.005)
Hours Restrictions				
Part-Time Restrictions	-1.647***	0.370***	0.046	0.116*
	(0.082)	(0.084)	(0.069)	(0.058)
Fixed Costs	-1.093***	0.448***	0.112	0.181***
	(0.070)	(0.085)	(0.076)	(0.063)
Observed Heterogeneity				
In β_C Only	-0.335***	0.046	0.010	0.013
	(0.057)	(0.035)	(0.023)	(0.022)
In β_L Only	-0.381***	0.136***	-0.034	-0.001
	(0.061)	(0.037)	(0.022)	(0.023)
In β_C And β_L	-0.475***	0.130***	-0.103***	-0.068***
	(0.070)	(0.044)	(0.025)	(0.024)
Unobserved Heterogeneity				
In β_C Only	0.005	-0.020	-0.098**	-0.088**
	(0.014)	(0.027)	(0.038)	(0.036)
In β_L Only	0.005	-0.077***	-0.010	-0.017
	(0.013)	(0.026)	(0.038)	(0.036)
In β_C And β_L	-0.041***	-0.052**	-0.096***	-0.088***
	(0.013)	(0.025)	(0.030)	(0.029)
In β_C And β_L With Correlation	-0.119***	-0.107**	-0.132***	-0.122***
	(0.016)	(0.040)	(0.036)	(0.036)
Wage Imputation				
Full Sample, No Correction	-0.811***	2.169***	2.052***	2.110***
	(0.119)	(0.098)	(0.123)	(0.112)
Full Sample, Error Integrated Out	-0.530***	1.278***	1.390***	1.397***
	(0.048)	(0.140)	(0.106)	(0.103)
Full Sample, 1 Random Draw	-0.104**	0.060	0.227	0.199
	(0.049)	(0.053)	(0.140)	(0.127)
Non-Workers, Error Integrated Out	0.000	0.042	0.013	0.021
	(0.067)	(0.050)	(0.035)	(0.032)
Non-Workers, 1 Random Draw	0.070	-0.187***	-0.265***	-0.266***
	(0.056)	(0.033)	(0.047)	(0.042)
Constant	1.004***	-0.940***	-0.529***	-0.623***
	(0.121)	(0.114)	(0.104)	(0.096)
Labor Supply Type Fixed Effects	Yes	Yes	Yes	Yes
Observations	16,730	13,219	13,219	13,219
Explanatory Power R^2	0.854	0.832	0.769	0.806

Notes: Uncompensated labor supply elasticities are simulated by increasing the individual wage rates by 10 percent and taking the median individual labor supply response. Dependent variables have been standardized, i.e., an estimate of 1.0 indicates an increase of one standard deviation in the outcome. The AIC is negatively related to the statistical fit of the model – the better the fit, the lower the AIC. Standard errors clustered by labor supply group and wage imputation procedure (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

Bibliography

Aaberge, R., J. K. Dagsvik, and S. Strøm (1995). "Labor Supply Responses and Welfare Effects of Tax Reforms". *Scandinavian Journal of Economics* 97.4, pp. 635–659.

Aaberge, R., U. Colombino, and S. Strøm (1999). "Labour Supply in Italy: An Empirical Analysis of Joint Houshold Decisions, with Taxes and Quantity Constraints". *Journal of Applied Econometrics* 14.4, pp. 403–422.

Aaberge, R., U. Colombino, and T. Wennemo (2009). "Evaluating Alternative Representations of the Choice Sets in Models of Labor Supply". *Journal of Economic Surveys* 23.3, pp. 586–612.

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011). "Reference Points and Effort Provision". *American Economic Review* 101.2, pp. 470–492.

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). "The Colonial Origins of Comparative Development: An Empirical Investigation". *American Economic Review* 91.5, pp. 1369–1401.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015). "The Economics of Density: Evidence From the Berlin Wall". *Econometrica* 83.6, pp. 2127–2189.

Ahrens, S., I. Pirschel, and D. J. Snower (2014). "A Theory of Wage Adjustment under Loss Aversion". *IZA Discussion Paper* 8699.

Albats, Y. (1995). *KGB: State Within a State*. London: I.B.Tauris & Co Ltd.

Alda, H., S. Bender, and H. Gartner (2005). "The Linked Employer-Employee Dataset Created From the IAB Establishment Panel and the Process-Produced Data of the IAB (LIAB)". *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 125.2, pp. 327–336.

Alesina, A. and N. Fuchs-Schündeln (2007). "Goodbye Lenin (or Not?): The Effect of Communism on People's Preferences". *American Economic Review* 97.4, pp. 1507–1528.

Alesina, A. and P. Giuliano (2015). "Culture and Institutions". *Journal of Economic Literature* 53.4, pp. 898–944.

Alesina, A., P. Giuliano, and N. Nunn (2013). "On the Origins of Gender Roles: Women and the Plough". *Quarterly Journal of Economics* 128.2, pp. 469–530.

Algan, Y. and P. Cahuc (2010). "Inherited Trust and Growth". *American Economic Review* 100.5, pp. 2060–2092.

– (2014). "Trust, Growth, and Well-Being: New Evidence and Policy Implications". In: *Handbook of Economic Growth*. Ed. by P. Aghion and S. Durlauf. Vol. 2. North Holland, Elsevier. Chap. 2.

Arendt, H. (1951). *The Origins of Totalitarianism.* New York: Harcourt, Brace and Company.

Arrow, K. J. (1972). “Gifts and Exchanges”. *Philosophy & Public Affairs* 1.4, pp. 343–362.

Bargain, O. and A. Peichl (2016). “Own-Wage Labor Supply Elasticities: Variation across Time and Estimation Methods”. *IZA Journal of Labor Economics* 5.10, pp. 1–31.

Bargain, O., K. Orsini, and A. Peichl (2014). “Comparing Labor Supply Elasticities in Europe and the US: New Results”. *Journal of Human Resources* 49.3, pp. 723–838.

Becker, S., K. Boeckh, C. Hainz, and L. Wößmann (2016). “The Empire Is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy”. *Economic Journal* 126.590, pp. 40–74.

Becker, S. O. and L. Wößmann (2009). “Was Weber Wrong? A Human Capital Theory of Protestant Economic History”. *Quarterly Journal of Economics* 124.2, pp. 531–596.

Benzarti, Y., D. Carlon, J. Harju, and T. Kosonen (2017). “What Goes Up May Not Come Down: Asymmetric Incidence of Value Added Taxes”. *NBER Working Paper* 23849.

Berg, J., J. Dickhaut, and K. McCabe (1995). “Trust, Reciprocity and Social History”. *Games and Economic Behavior* 10.1, pp. 122–142.

Bewley, T. F. (2005). “Fairness, Reciprocity, and Wage Rigidity”. In: *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life.* Ed. by H. Gintis, S.Bowles, R. T. Boyd, and E. Fehr. MIT Press.

Bisin, A. and T. Verdier (2000). “‘Beyond the Melting Pot’: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits”. *Quarterly Journal of Economics* 115.3, pp. 955–988.

Bisin, A., G. Topa, and T. Verdier (2004). “Religious Intermarriage and Socialization in the United States”. *Journal of Political Economy* 112.3, pp. 615–664.

Bloemen, H. G. and A. Kapteyn (2008). “The Estimation of Utility-Consistent Labor Supply Models By Means of Simulated Scores”. *Journal of Applied Econometrics* 23.4, pp. 395–422.

Blundell, R. and T. E. MaCurdy (1999). “Labor Supply: A Review of Alternative Approaches”. In: *Handbook of Labor Economics.* Ed. by O. Ashenfelter and D. Card. Vol. 3. Amsterdam: North-Holland. Chap. 27, pp. 1559–1695.

Blundell, R. and A. Shephard (2012). “Employment, Hours of Work and the Optimal Taxation of Low-Income Families”. *Review of Economic Studies* 79.2, pp. 481–510.

Blundell, R., A. Duncan, J. McCrae, and C. Meghir (1999). “Evaluating In-Work Benefit Reform: the Working Families Tax Credit in the UK”. Mimeo, presented at the Joint Center for Poverty Research Conference, Northwestern University.

– (2000). “The Labour Market Impact of the Working Families’ Tax Credit”. *Fiscal Studies* 21.1, pp. 75–104.

Borjas, G. J. (1980). "The Relationship between Wages and Weekly Hours of Work: The Role of Division Bias". *Journal of Human Resources* 15.3, pp. 409–423.

Bradley, S. (2017). "Inattention to Deferred Increases in Tax Bases: How Michigan Homebuyers are Paying for Assessment Limits". *Review of Economics and Statistics* 99.1, pp. 53–66.

Branigan, T. (2010). *Chinese Police Chief Boasts of Recruiting One in 33 Residents as Informants*. The Guardian, Februrary 10, 2010, `https://www.theguardian.com/world/2010/feb/10/china-police-informants-surveillance`.

Breunig, R., D. A. Cobb-Clark, and X. Gong (2008). "Improving the Modelling of Couples' Labour Supply". *Economic Record* 84.267, pp. 466–485.

Bruce, G. (2010). *The Firm – The Inside Story of the Stasi*. New York: Oxford University Press.

Brueckner, J. K. (1981). "Labor Mobility and the Incidence of the Residential Property Tax". *Journal of Urban Economics* 10.2, pp. 173–182.

Buhrmester, M., T. Kwang, and S. D. Gosling (2011). "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6.1, pp. 3–5.

Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler (1997). "Labor Supply of New York City Cabdrivers: One Day at a Time". *Quarterly Journal of Economics* 112.2, pp. 407–441.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). "Bootstrap-Based Improvements for Inference with Clustered Errors". *Review of Economics and Statistics* 90.3, pp. 414–427.

Capon, F. (2015). *"Soviet-Era Snitching" Returns to Russia*. Newsweek, June 1, 2015, `http://europe.newsweek.com/soviet-era-snitching-returns-russia-328036`.

Card, D. and G. B. Dahl (2011). "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior". *Quarterly Journal of Economics* 126.1, pp. 103–143.

Carroll, R. J. and J. Yinger (1994). "Is The Property Tax a Benefit Tax? The Case of Rental Housing". *National Tax Journal* 47.2, pp. 295–316.

Charness, G. and P. Kuhn (2011). "Lab Labor: What Can Labor Economists Learn From the Lab?" In: *Handbook of Labor Economics Vol 4A*. Ed. by O. Ashenfelter and D. Card. Amsterdam: North Holland, pp. 229–330.

Chemin, M. and A. Kurmann (2014). "Do Workers Feel Entitled to High Wages? Evidence from a Field Experiment". *Mimeo*.

Chetty, R., A. Looney, and K. Kroft (2009). "Salience and Taxation: Theory and Evidence". *American Economic Review* 99.4, pp. 1145–1177.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011). "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins". *American Economic Review: Papers and Proceedings* 101.3, pp. 471–475.

Childs, D. and R. Popplewell (1996). *The Stasi: The East German Intelligence and Security Service*. New York: New York University Press.

Cohn, A., E. Fehr, and L. Goette (2015). "Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment". *Management Science* 61.8, pp. 1777–1794.

Combes, P.-P., G. Duranton, and L. Gobillon (2016). "The Production Function For Housing: Evidence From France". *CEPR Discussion Paper* 11669.

Crawford, V. P. and J. Meng (2011). "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income". *American Economic Review* 101.5, pp. 1912–1932.

Dagsvik, J. K., Z. Jia, T. Kornstadt, and T. O. Thoresen (2014). "Theoretical and Practical Arguments For Modeling Labor Supply as a Choice Among Latent Jobs". *Journal of Economic Surveys* 28.1, pp. 134–151.

Dagsvik, J. K. and S. Strøm (2004). "Sectoral Labor Supply, Choice Restrictions and Functional Form". *Statistics Norway Discussion Paper* 388.

– (2006). "Sectoral Labor Supply, Choice Restrictions and Functional Form". *Journal of Applied Econometrics* 21.6, pp. 803–826.

Dagsvik, J. K., Z. Jia, K. Orsini, and G. van Camp (2011). "Subsidies on Low-Skilled Workers' Social Security Contributions: The Case of Belgium". *Empirical Economics* 40.3, pp. 779–806.

de Bartolomé, C. A. M. and S. S. Rosenthal (1999). "Property Tax Capitalization in a Model with Tax-Deferred Assets, Standard Deductions, and the Taxation of Nominal Interest". *Review of Economics and Statistics* 81.1, pp. 85–95.

DellaVigna, S. and D. Pope (2018). "What Motivates Effort? Evidence and Expert Forecasts". *Review of Economic Studies* (forthcoming).

DellaVigna, S., A. Lindner, B. Reizer, and J. F. Schmieder (2018). "Reference-Dependent Job Search: Evidence from Hungary". *Quarterly Journal of Economics* (forthcoming).

Diamond, P. and E. Saez (2011). "The Case for a Progressive Tax: From Basic Research to Policy Recommendations". *Journal of Economic Perspectives* 25.4, pp. 165–190.

Diamond, R. (2016). "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000". *American Economic Review* 106.3, pp. 479–524.

Dixit, A. K. and J. E. Stiglitz (1977). "Monopolistic Competition and Optimum Product Diversity". *American Economic Review* 67.3, pp. 297–308.

Doerrenberg, P., D. Duncan, and M. Löffler (2016). "Asymmetric Labor-Supply Responses to Wage-Rate Changes: Evidence from a Field Experiment". *IZA Discussion Paper* 9683.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2008). "Representative Trust and Reciprocity: Prevalence and Determinants". *Economic Inquiry* 46.1, pp. 84–90.

– (2009). "Homo Reciprocans: Survey Evidence On Behavioural Outcomes". *Economic Journal* 119.3, pp. 592–612.

– (2012). “The Intergenerational Transmission of Risk and Trust Attitudes”. *Review of Economic Studies* 79.2, pp. 645–677.

Dube, A., T. W. Lester, and M. Reich (2010). “Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties”. *Review of Economics and Statistics* 92.4, pp. 945–964.

Dunn, L. F. (1996). “Loss Aversion and Adaptation in the Labor Market: Empirical Indifference Functions and Labor Supply”. *Review of Economics and Statistics* 78.3, pp. 441–450.

Dusansky, R., M. Ingber, and N. Karatjas (1981). “The Impact of Property Taxation on Housing Values and Rents”. *Journal of Urban Economics* 10.2, pp. 240–255.

Eissa, N. and J. B. Liebman (1996). “Labor Supply Response to the Earned Income Tax Credit”. *Quarterly Journal of Economics* 111.2, pp. 605–637.

Eissa, N., H. J. Kleven, and C. T. Kreiner (2008). “Evaluation of Four Tax Reforms in The United States: Labor Supply and Welfare Effects For Single Mothers”. *Journal of Public Economics* 92.3-4, pp. 795–816.

Eklöf, M. and H. Sacklén (2000). “The Hausman-MaCurdy Controversy: Why Do the Results Differ across Studies?” *Journal of Human Resources* 35.1, pp. 204–220.

Engelmann, R. and S. Schumann (1995). “Der Ausbau des Überwachungsstaates. Der Konflikt Ulbricht-Wollweber und die Neuausrichtung des Staatssicherheitsdienstes der DDR 1957”. *Vierteljahreshefte zur Zeitgeschichte* 43.2, pp. 341–378.

England, R. W. (2016). “Tax Incidence and Rental Housing: A Survey and Critique of Research”. *National Tax Journal* 69.2, pp. 435–460.

Epple, D., B. Gordon, and H. Sieg (2010). “A New Approach to Estimating the Production Function for Housing”. *American Economic Review* 100.3, pp. 905–924.

Ericson, K. M. M. and A. Fuster (2011). “Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments”. *Quarterly Journal of Economics* 126.4, pp. 1879–1907.

Ericson, P. and L. Flood (1997). “A Monte Carlo Evaluation of Labor Supply Models”. *Empirical Economics* 22.3, pp. 431–460.

Euwals, R. and A. van Soest (1999). “Desired and Actual Labour Supply of Unmarried Men and Women in the Netherlands”. *Labour Economics* 6.1, pp. 95–118.

Evers, M., R. de Mooij, and D. van Vuuren (2008). “The Wage Elasticity of Labour Supply: A Synthesis of Empirical Estimates”. *De Economist* 156.1, pp. 25–43.

Fajgelbaum, P., E. Morales, J. C. Suárez Serrato, and O. Zidar (2016). “State Taxes and Spatial Misallocation”. Mimeo.

Falk, A., E. Fehr, and C. Zehnder (2006). “Fairness Perceptions and Reservation Wages – the Behavioral Effects of Minimum Wage Laws”. *Quarterly Journal of Economics* 121.4, pp. 1347–1381.

Falter, J. W. and D. Hänisch (1990). "Election and Social Data of the Districts and Municipalities of the German Empire from 1920 to 1933". *GESIS Data Archive, Cologne* ZA8013. DOI: 10.4232/1.8013.

Fehr, E. and L. Goette (2007). "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment". *American Economic Review* 97.1, pp. 298–317.

Ferreira, F. (2010). "You Can Take It With You: Proposition 13 Tax Benefits, Residential Mobility, and Willingness to Pay for Housing Amenities". *Journal of Public Economics* 94.9–10, pp. 661–673.

Flood, L., J. Hansen, and R. Wahlberg (2004). "Household Labor Supply and Welfare Participation in Sweden". *Journal of Human Resources* 39.4, pp. 1008–1032.

Flood, L., R. Wahlberg, and E. Pylkkänen (2007). "From Welfare to Work: Evaluating a Tax and Benefit Reform Targeted at Single Mothers in Sweden". *LABOUR* 21.3, pp. 443–471.

Fontana, N., T. Nannicini, and G. Tabellini (2017). "Historical Roots of Political Extremism: The Effects of Nazi Occupation of Italy". *IGIER Working Paper* 618.

Fouka, V. and H.-J. Voth (2016). "Reprisals Remembered: German-Greek Conflict and Car Sales during the Euro Crisis". Mimeo.

Friehe, T., M. Pannenberg, and M. Wedow (2015). "Let Bygones be Bygones? Socialist Regimes and Personalities in Germany". *SOEPpapers* 776.

Fuchs-Schündeln, N. and T. A. Hassan (2016). "Natural Experiments in Macroeconomics". In: *Handbook of Macroeconomics*. Ed. by J. B. Taylor and H. Uhlig. Vol. 2. North Holland, Elsevier. Chap. 12.

Fuchs-Schündeln, N. and P. Masella (2016). "Long-Lasting Effects of Socialist Education". *Review of Economics and Statistics* 98.3, pp. 428–441.

Fuest, C., A. Peichl, and S. Siegloch (2018). "Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany". *American Economic Review* 108.2, pp. 393–418.

Fukuyama, F. (1995). *Trust*. New York: Free Press.

Fulbrook, M. (1995). *Anatomy of a Dictatorship: Inside the GDR 1949–1989*. Oxford: Oxford University Press.

– (2005). *The People's State. East German Society from Hitler to Honecker*. New Haven: Yale University Press.

– (2009). *A History of Germany 1918–2008: The Divided Nation*. 3rd ed. Malden, Oxford, Chichester: Wiley-Blackwell.

Gieseke, J. (2014). *The History of the Stasi: East Germany's Secret Police, 1945–1990*. New York, Oxford: Berghahn Books.

Gill, D. and U. Schröter (1991). *Das Ministerium für Staatssicherheit. Anatomie des Mielke-Imperiums*. Berlin: Rowohlt.

Glaeser, E. L. (2009). “The Death and Life of Cities”. In: *Making Cities Work: Prospects and Policies for Urban America.* Ed. by R. P. Inman. Princeton University Press, pp. 22–62.

Glaeser, E. L., D. I. Laibson, J. A. Scheinkman, and C. L. Soutter (2000). “Measuring Trust”. *Quarterly Journal of Economics* 115.3, pp. 811–846.

Glitz, A. and E. Meyersson (2016). “Industrial Espionage and Productivity”. Mimeo.

Gneezy, U. and J. A. List (2006). “Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments”. *Econometrica* 74.5, pp. 1365–1384.

Goette, L., D. Huffman, and E. Fehr (2004). “Loss Aversion and Labor Supply”. *Journal of the European Economic Association* 2.2-3, pp. 216–228.

Goodchild, M. F. and N. S.-N. Lam (1980). “Areal Interpolation: A Variant of the Traditional Spatial Problem”. *Geo-Processing* 1, pp. 297–312.

Grashoff, U. (2011). *Schwarzwohnen. Die Unterwanderung der staatlichen Wohnraumlenkung in der DDR.* Göttingen: V&R unipress.

Guiso, L., P. Sapienza, and L. Zingales (2004). “The Role of Social Capital in Financial Development”. *American Economic Review* 94.3, pp. 526–556.

– (2006). “Does Culture Affect Economic Outcomes?” *Journal of Economic Perspectives* 20.2, pp. 23–48.

– (2008). “Social Capital as Good Culture”. *Journal of the European Economic Association* 6.2–3, pp. 295–320.

– (2009). “Cultural Biases in Economic Exchange?” *Quarterly Journal of Economics* 124.3, pp. 1095–1131.

Haan, P. (2006). “Much Ado About Nothing: Conditional Logit vs. Random Coefficient Models For Estimating Labour Supply Elasticities”. *Applied Economics Letters* 13.4, pp. 251–256.

Hall, R. E. and C. I. Jones (1999). “Why do Some Countries Produce So Much More Output Per Worker than Others?” *Quarterly Journal of Economics* 114.1, pp. 83–116.

Hamilton, B. W. (1975). “Zoning and Property Taxation in a System of Local Governments”. *Urban Studies* 12.2, pp. 205–211.

– (1976). “Capitalization of Intrajurisdictional Differences in Local Tax Prices”. *American Economic Review* 66.5, pp. 743–753.

Harrison, M. and I. Zaksauskiene (2015). “Counter-Intelligence in a Command Economy”. *Economic History Review* 69.1, pp. 131–158.

Hausman, J. A. (1981). “Labor Supply”. In: *How Taxes Affect Economic Behavior.* Ed. by H. J. Aaron and J. A. Pechman. Washington, D.C.: Brookings Institution, pp. 27–72.

Heckman, J. J. (1993). “What Has Been Learned About Labor Supply in the Past Twenty Years?” *American Economic Review: Papers and Proceedings* 83.2, pp. 116–121.

Heinberg, J. D. and W. E. Oates (1970). "The Incidence of Differential Property Taxes on Urban Housing: A Comment and Some Further Evidence". *National Tax Journal* 23.1, pp. 92–98.

Hennig-Schmidt, H., A. Sadrieh, and B. Rockenbach (2010). "In Search of Workers' Real Effort Reciprocity – a Field and a Laboratory Experiment". *Journal of the European Economic Association* 8.4, pp. 817–837.

Hilber, C. A. L. and W. Vermeulen (2016). "The Impact of Supply Constraints on House Prices in England". *Economic Journal* 126.591, pp. 358–405.

Holmes, T. J. (1998). "The Effect of State Policy on the Location of Manufacturing: Evidence from State Borders". *Journal of Political Economy* 106.4, pp. 667–705.

Horsch, H. (1997). "'Hat nicht wenigstens die Stasi die Stimmung im Lande gekannt?' MfS und SED im Bezirk Karl-Marx-Stadt. Die Entmachtung der Staatssicherheit in den Regionen, Teil 3". *BF informiert* 19/1997.

Horton, J. J. and L. B. Chilton (2010). "The Labor Economics of Paid Crowdsourcing". *Proceedings of the 11th ACM Conference on Electronic Commerce.*

Horton, J. J., D. G. Rand, and R. J. Zeckhauser (2011). "The Online Laboratory: Conducting Experiments in a Real Labor Market". *Experimental Economics* 14, pp. 399–425.

Hoynes, H. W. (1996). "Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation Under AFDC-UP". *Econometrica* 64.2, pp. 295–332.

Hyman, D. N. and E. C. J. Pasour (1973). "Property Tax Differentials and Residential Rents in North Carolina". *National Tax Journal* 26.2, pp. 303–307.

Immervoll, H., H. J. Kleven, C. T. Kreiner, and N. Verdelin (2011). "Optimal Tax and Transfer Programs For Couples With Extensive Labor Supply Responses". *Journal of Public Economics* 95.11, pp. 1485–1500.

infas. *DDR-Kreisstrukturdaten.* München: infas-Absatzplanung.

Jacob, M. and M. Tyrell (2010). "The Legacy of Surveillance: An Explanation for Social Capital Erosion and the Persistent Economic Disparity between East and West Germany". Mimeo, presented at the Sciences Po/IZA Workshop on Trust, Civic Spirit and Economic Performance in June 2010.

Jacobs, A. and J. Ansfield (2011). *Well-Oiled Security Apparatus in China Stifles Calls for Change.* The New York Times, February 28, 2011, `http://www.nytimes.com/2011/03/01/world/asia/01china.html`.

Kahn, S. (1997). "Evidence of Nominal Wage Stickiness from Microdata". *American Economic Review* 87.5, pp. 993–1008.

Kahneman, D. and A. Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica* 47.2, pp. 263–292.

Kaur, S. (2018). “Nominal Wage Rigidity in Village Labor Markets”. *American Economic Review* (forthcoming).

Keane, M. P. and R. Rogerson (2012). “Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom”. *Journal of Economic Literature* 50.2, pp. 464–476.

Keane, M. P. (2010). “Structural vs. Atheoretic Approaches to Econometrics”. *Journal of Econometrics* 156.1, pp. 3–20.

– (2011). “Labor Supply and Taxes: A Survey”. *Journal of Economic Literature* 49.4, pp. 961–1075.

Keane, M. P. and R. Moffitt (1998). “A Structural Model of Multiple Welfare Program Participation and Labor Supply”. *International Economic Review* 39.3, pp. 553–589.

Keane, M. P. and N. Wasi (2012). “Comparing Alternative Models of Heterogeneity in Consumer Choice Behavior”. *Journal of Applied Econometrics* 28.6, pp. 1018–1045.

Kessler, J. B. and M. I. Norton (2016). “Tax Aversion in Labor Supply”. *Journal of Economic Behavior and Organization* 124, pp. 15–28.

King, G., O. Rosen, M. Tanner, and A. F. Wagner (2008). “Ordinary Economic Voting Behavior in the Extraordinary Election of Adolf Hitler”. *Journal of Economic History* 68.4, pp. 951–996.

Kline, P. and E. Moretti (2014). “People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs”. *Annual Review of Economics* 6.1, pp. 629–662.

Knabe, H. (1999). “Die feinen Waffen der Diktatur. Nicht-strafrechtliche Formen politischer Verfolgung in der DDR”. In: *Die DDR – Erinnerungen an einen untergegangenen Staat.* Ed. by H. Timmermann. Vol. 88. Dokumente und Schriften der Europäischen Akademie Otzenhausen. Berlin: Duncker & Humboldt, pp. 191–220.

Knack, S. and P. Keefer (1997). “Does Social Capital Have an Economic Payoff? A Cross-Country Investigation”. *Quarterly Journal of Economics* 112.4, pp. 1251–1288.

Kowalczuk, I.-S. (2013). *Stasi Konkret. Überwachung und Repression in der DDR.* Munich: C.H.Beck.

Kőszegi, B. and M. Rabin (2006). “A Model of Reference-Dependent Preferences”. *Quarterly Journal of Economics* 121.4, pp. 1133–1165.

– (2007). “Reference-Dependent Risk Attitudes”. *American Economic Review* 97.4, pp. 1047–1073.

– (2009). “Reference-Dependent Consumption Plans”. *American Economic Review* 99.3, pp. 909–936.

Kube, S., M. A. Marechal, and C. Puppe (2013). “Do Wage Cuts Damage Work Morale? Evidence From a Natural Field Experiment”. *Journal of the European Economic Association* 11.4, pp. 853–870.

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015). “How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments”. *American Economic Review* 105.4, pp. 1478–1508.

La Porta, R., F. López de Silanes, A. Shleifer, and R. W. Vishny (1997). “Trust in Large Organizations”. *American Economic Review: Papers and Proceedings* 87.2, pp. 333–338.

– (1998). “Law and Finance”. *Journal of Political Economy* 106.6, pp. 1113–1155.

Lee, D. S. and T. Lemieux (2010). “Regression Discontinuity Designs in Economics”. *Journal of Economic Literature* 48.2, pp. 281–355.

Lichter, A., M. Löffler, and S. Siegloch (2016). “The Long-Term Costs of Government Surveillance: Insights from Stasi Spying in East Germany”. *CESifo Working Paper* 6042.

Löffler, M. (2013). “Fitting Complex Mixed Logit Models: An Application to Labor Supply”. Mimeo, presented at the Stata Conference in New Orleans, July 2013.

Löffler, M., A. Peichl, N. Pestel, S. Siegloch, and E. Sommer (2014a). “Documentation IZAΨMOD v3.0: The IZA Policy Simulation Model”. *IZA Discussion Paper* 8553.

Löffler, M., A. Peichl, and S. Siegloch (2014b). “Structural Labor Supply Models and Wage Exogeneity”. *IZA Discussion Paper* 8281.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Lutz, B. (2015). “Quasi-Experimental Evidence on the Connection between Property Taxes and Residential Capital Investment”. *American Economic Journal: Economic Policy* 7.1, pp. 300–330.

Lyytikäinen, T. (2009). “Three-Rate Property Taxation and Housing Construction”. *Journal of Urban Economics* 65.3, pp. 305–313.

MaCurdy, T. E., D. Green, and H. Paarsch (1990). “Assessing Empirical Approaches for Analyzing Taxes and Labor Supply”. *Journal of Human Resources* 25.3, pp. 412–490.

Magruder, J. R. (2012). “High Unemployment Yet Few Small Firms: The Role of Centralized Bargaining in South Africa”. *American Economic Journal: Applied Economics* 4.3, pp. 138–166.

Mann, H. B. and D. R. Whitney (1947). “On a Test Whether One of Two Random Variables Is Stochastically Larger than the Other”. *Annals of Mathematical Statistics* 18, pp. 50–60.

Manski, C. F. (2014). “Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy”. *Quantitative Economics* 5.1, pp. 145–174.

Mason, W. and S. Suri (2011). “Conducting Behavioral Research on Amazon’s Mechanical Turk”. *Behavioural Research* 44, pp. 1–23.

Mauro, P. (1995). “Corruption and Growth”. *Quarterly Journal of Economics* 110.3, pp. 681–712.

Max Planck Institute for Demographic Research (MPIDR) and Chair for Geodesy and Geoinformatics, University of Rostock (CGG) (2011). *MPIDR Population History GIS Collection*. (Partly based on Hubatsch, W. and T. Klein (eds) 1975 ff.: Grundriß der deutschen Verwaltungsgeschichte, Marburg, and Bundesamt für Kartographie und Geodäsie: VG 2500 Verwaltungsgebiete [Ebenen] 1:2.500.000, Stand 01.01.2009, Frankfurt am Main). Rostock.

McCrary, J. (2007). “The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police”. *American Economic Review* 97.1, pp. 318–353.

McFadden, D. (1974). “Conditional Logit Analysis of Qualitative Choice Behavior”. In: *Frontiers in Econometrics*. Ed. by P. Zarembka. New York: Academic Press. Chap. 4, pp. 105–142.

Meghir, C. and D. Phillips (2010). “Labour Supply and Taxes”. In: *Dimensions of Tax Design: The Mirrlees Review*. Ed. by J. Mirrless, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba. Oxford University Press, pp. 202–274.

Melitz, M. J. (2003). “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”. *Econometrica* 71.6, pp. 1695–1725.

Michaillat, P. and E. Saez (2017). “Optimal Public Expenditure with Inefficient Unemployment”. Mimeo.

Mieszkowski, P. (1972). “The Property Tax: An Excise Tax Or a Profits Tax?” *Journal of Public Economics* 1.1, pp. 73–96.

Mieszkowski, P. and G. R. Zodrow (1989). “Taxation and The Tiebout Model: The Differential Effects of Head Taxes, Taxes on Land Rents, and Property Taxes”. *Journal of Economic Literature* 27.3, pp. 1098–1146.

Moffitt, R. (1983). “An Economic Model of Welfare Stigma”. *American Economic Review* 73.5, pp. 1023–1035.

Moretti, E. (2011). “Local Labor Markets”. In: *Handbook of Labor Economics*. Ed. by O. Ashenfelter and D. Card. Vol. 4B. North-Holland, pp. 1237–1313.

Müller-Enbergs, H. (1996). *Inoffizielle Mitarbeiter des Ministeriums für Staatssicherheit. Richtlinien und Durchführungsbestimmungen*. Berlin: Ch. Links Verlag.

– (2008). *Inoffizielle Mitarbeiter des Ministeriums für Staatssicherheit. Teil 3: Statistik*. Berlin: Ch. Links Verlag.

Naimark, N. M. (1994). “‘To Know Everything and to Report Everything Worth Knowing’: Building the East German Police State, 1945–1949”. *Cold War International History Project Working Paper* 10.

Niemann, M. (2007). *Die Sekretäre der SED-Bezirksleitungen 1952 bis 1989*. Paderborn: Ferdinand Schöningh.

Nunn, N. (2008). “The Long-Term Effects of Africa’s Slave Trades”. *Quarterly Journal of Economics* 123.1, pp. 139–176.

Nunn, N. and L. Wantchekon (2011). “The Slave Trade and the Origins of Mistrust in Africa”. *American Economic Review* 101.7, pp. 3221–3252.

Oates, W. E. and W. A. Fischel (2016). “Are Local Property Taxes Regressive, Progressive, Or What?” *National Tax Journal* 69.2, pp. 415–434.

Orr, L. L. (1968). “The Incidence of Differential Property Taxes on Urban Housing”. *National Tax Journal* 21.3, pp. 253–262.

Orr, L. L. (1970). "The Incidence of Differential Property Taxes: A Response". *National Tax Journal* 23.1, pp. 99–101.

– (1972). "The Incidence of Differential Property Taxes on Urban Housing: Reply". *National Tax Journal* 25.2, pp. 217–220.

Palmon, O. and B. A. Smith (1998). "New Evidence on Property Tax Capitalization". *Journal of Political Economy* 106.5, pp. 1099–1111.

Paolacci, G., J. Chandler, and P. G. Ipeirotis (2010). "Running Experiments on Amazon Mechanical Turk". *Judgment and Decision Making* 5.5.

Pew Research Center (2014). *Spring Political Typology Survey*. `http://www.pewresearch.org/fact-tank/2015/05/29/what-americans-think-about-nsa-surveillance-national-security-and-privacy/`.

Plassmann, F. and T. N. Tideman (2000). "A Markov Chain Monte Carlo Analysis of the Effect of Two-Rate Property Taxes on Construction". *Journal of Urban Economics* 47.2, pp. 216–247.

Pope, D. G. and M. E. Schweitzer (2011). "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes". *American Economic Review* 101.1, pp. 129–157.

Popplewell, R. (1992). "The Stasi and the East German Revolution of 1989". *Contemporary European History* 1.1, pp. 37–63.

Post, T., M. J. van den Assem, G. Baltussen, and R. H. Thaler (2008). "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show". *American Economic Review* 98.1, pp. 38–71.

Poterba, J. M. (1984). "Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach". *Quarterly Journal of Economics* 99.4, pp. 729–752.

Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

– (1995). "Bowling Alone: America's Declining Social Capital". *Journal of Democracy* 6.1, pp. 65–78.

– (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Reich, J. (1997). "Sicherheit und Feigheit – der Käfer im Brennglas". In: *Staatspartei und Staatssicherheit. Zum Verhältnis von SED und MfS*. Ed. by S. Suckut. Berlin: Ch. Links Verlag, pp. 25–37.

Roback, J. (1982). "Wages, Rents, and the Quality of Life". *Journal of Political Economy* 90.6, pp. 1257–1278.

Rodrik, D., A. Subramanian, and F. Trebbi (2004). "Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development". *Journal of Economic Growth* 9.2, pp. 131–165.

Rosen, S. (1979). “Wage-Based Indexes of Urban Quality of Life”. In: *Current Issues in Urban Economics*. Ed. by P. Mieszkowski and M. Straszheim. Baltimore: Johns Hopkins University Press, pp. 74–104.

Rothstein, J. (2010). “Is the EITC as Good as an NIT? Conditional Cash Transfers and Tax Incidence”. *American Economic Journal: Economic Policy* 2.1, pp. 177–208.

Rudolph, H. (1990). “Beschäftigungsstrukturen in der DDR vor der Wende”. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 4, pp. 474–503.

Saez, E., J. Slemrod, and S. H. Giertz (2012). “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review”. *Journal of Economic Literature* 50.1, pp. 3–50.

Saiz, A. (2010). “The Geographic Determinants of Housing Supply”. *Quarterly Journal of Economics* 125.3, pp. 1253–1296.

Schneier, B. (2013). *NSA Secrets Kill Our Trust*. `http://edition.cnn.com/2013/07/31/opinion/schneier-nsa-trust/`.

Shan, H. (2010). “Property Taxes and Elderly Mobility”. *Journal of Urban Economics* 67.6, pp. 194–205.

Shleifer, A. (1997). “Government in Transition”. *European Economic Review* 41.3–5, pp. 385–410.

Simon, D. (2016). “Does Early Life Exposure to Cigarette Smoke Permanently Harm Childhood Welfare? Evidence from Cigarette Tax Hikes”. *American Economic Journal: Applied Economics* 8.4, pp. 128–159.

Smith, A. (1776). *An Inquiry Into the Nature and Causes of the Wealth of Nations*. Ed. by E. Cannan. University of Chicago Press.

Smith, M., D. Yagan, O. Zidar, and E. Zwick (2017). “Capitalists in the Twenty-First Century”. Mimeo.

Socio-Economic Panel (SOEP) (2015). *Data for Years 1984–2014, Version 31*. DOI: 10.5684/soep.v31.

Spahn, P. B. (2004). “Land Taxation in Germany”. In: *International Handbook of Land and Property Taxation*. Ed. by R. M. Bird and N. E. Slack. Edward Elgar Publishing, pp. 98–106.

Statistisches Bundesamt (1993). “Bevölkerungsstatistische Übersichten 1946 bis 1989”. *Sonderreihe mit Beiträgen für das Gebiet der ehemaligen DDR* 3, p. 149.

Suárez Serrato, J. C. and O. Zidar (2016). “Who Benefits from State Corporate Tax Cuts? A Local Labor Market Approach with Heterogeneous Firms”. *American Economic Review* 106.9, pp. 2582–2624.

Tabellini, G. (2008). “The Scope of Cooperation: Values and Incentives”. *Quarterly Journal of Economics* 123.3, pp. 905–950.

– (2010). “Culture and Institutions: Economic Development in the Regions of Europe”. *Journal of the European Economic Association* 8.4, pp. 677–716.

The Economist Intelligence Unit (2014). *Democracy Index*. `http://www.eui.com`.

Thorsnes, P. (1997). "Consistent Estimates of the Elasticity of Substitution between Land and Non-Land Inputs in the Production of Housing". *Journal of Urban Economics* 42.1, pp. 98–108.

Tiebout, C. M. (1956). "A Pure Theory of Local Expenditures". *Journal of Political Economy* 64.5, pp. 416–424.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge University Press.

van Soest, A. (1995). "Structural Models of Family Labor Supply – A Discrete Choice Approach". *Journal of Human Resources* 30.1, pp. 63–88.

van Soest, A. and M. Das (2001). "Family Labor Supply and Proposed Tax Reforms in the Netherlands". *De Economist* 149.2, pp. 191–218.

van Soest, A., M. Das, and X. Gong (2002). "A Structural Labour Supply Model With Flexible Preferences". *Journal of Econometrics* 107.1-2, pp. 345–374.

Voigtländer, N. and H.-J. Voth (2012). "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany". *Quarterly Journal of Economics* 127.3, pp. 1339–1392.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). "The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements". *Journal of Applied Social Science Studies* 127.1, pp. 139–169.

Wilcoxon, F. (1945). "Individual Comparisons By Ranking Methods". *Biometrics* 1, pp. 80–83.

Wolle, S. (2009). *Die heile Welt der Diktatur. Herrschaft und Alltag in der DDR 1971–1989*. Berlin: Ch. Links Verlag.

Yu, M. (2014). *Paranoia in Beijing With Million-Man Mobilization of Government Informants*. The Washington Times, October 23, 2014, `http://www.washingtontimes.com/news/2014/oct/23/inside-china-beijing-mobilizes-1-million-informant/`.

Zak, P. J. and S. Knack (2001). "Trust and Growth". *Economic Journal* 111.470, pp. 295–321.

Ziliak, J. P. and T. J. Kniesner (1999). "Estimating Life-Cycle Labor Supply Tax Effects". *Journal of Political Economy* 107.2, pp. 326–359.

Zodrow, G. R. (2001). "The Property Tax as a Capital Tax: A Room with Three Views". *National Tax Journal* 54.1, pp. 139–156.

Curriculum Vitae

Personal Details

Address	Max Löffler University of Cologne Center for Macroeconomic Research (CMR) Albertus-Magnus-Platz, D-50923 Cologne, Germany
Contact	+49 (0)221 470-8653, max.loeffler@wiso.uni-koeln.de
Born	January 24, 1988 in Halle (Saale)

Current and Past Positions

Since 06/2017	**Researcher and Teaching Assistant** Center for Macroeconomic Research, University of Cologne
04/2014–05/2017	**Researcher** Centre for European Economic Research (ZEW), Mannheim Research Group "International Distribution and Redistribution"
08/2012–03/2014	**Resident Research Affiliate** Institute for the Study of Labor (IZA), Bonn

Education

Since 10/2012	**Ph.D. Candidate in Economics** University of Cologne
10/2006–05/2012	**Diplom-Volkswirt** University of Cologne

Scholarships

08/2012–03/2014	**Scholarship for Ph.D. Students** Institute for the Study of Labor, Bonn
04/2009–05/2012	**Student Scholarship** Heinrich Böll Foundation, Berlin