Essays on the Economics of Local Labor Markets

by

Matthew J. Notowidigdo

B.S., Massachusetts Institute of Technology (2003)
M.Eng., Massachusetts Institute of Technology (2004)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2010

Author ....
Department of Economics
May 15, 2010

Certified by ....
David Autor
Professor of Economics
Thesis Supervisor

Certified by ....
Amy Finkelstein
Professor of Economics
Thesis Supervisor

Accepted by ....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation
and Development Economics
Chairman, Departmental Committee on Graduate Studies

Essays on the Economics of Local Labor Markets

by

Matthew J. Notowidigdo

Submitted to the Department of Economics
on May 15, 2010, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis studies the economics of local labor markets. There are three chapters in the thesis, and each chapter studies how economic outcomes are affected by local labor market conditions.

The first chapter studies the incidence of local labor demand shocks. This chapter starts from the observation that low-skill workers are comparatively immobile. When labor demand slumps in a city, college-educated workers tend to relocate whereas non-college workers are disproportionately likely to remain to face declining wages and employment. A standard explanation of these facts is that mobility is more costly for low-skill workers. This chapter proposes and tests an alternative explanation, which is that the incidence of adverse shocks is borne in large part by (falling) real estate rental prices and (rising) social transfers. These factors reduce the real cost of living differentially for low-income workers and thus compensate them, in part or in full, for declining labor demand. I develop a spatial equilibrium model which, appropriately parameterized, identifies both the magnitude of unobserved mobility costs by skill and the shape of the local housing supply curve. Nonlinear reduced form estimates using U.S. Census data document that positive labor demand shocks increase population more than negative shocks reduce population, that this asymmetry is larger for low-skill workers, and that such an asymmetry is absent for wages, housing values, and rental prices. Estimates of the full model using a nonlinear, simultaneous equations GMM estimator suggest that (1) the asymmetric population response is primarily accounted for by an asymmetric housing supply curve, (2) the differential migration response by skill is primarily accounted for by transfer payments, and (3) estimated mobility costs are at most modest and are comparable for high-skill and low-skill workers, suggesting that the primary explanation for the comparative immobility of low-skilled workers is not higher mobility costs per se, but rather a lower incidence of adverse labor demand shocks.

The second chapter, written jointly with Daron Acemoglu and Amy Finkelstein, studies how local area health spending responds to permanent changes in local area income. This chapter is motivated by the fact that health expenditures as a share of GDP have more than tripled over the last half century, and a common conjecture

is that this is primarily a consequence of rising real per capita income, which more than doubled over the same period. We investigate this hypothesis empirically by instrumenting for local area income with time-series variation in global oil prices between 1970 and 1990 interacted with cross-sectional variation in the oil reserves across different areas of the Southern United States. This strategy enables us to capture both the partial equilibrium and the local general equilibrium effects of an increase in income on health expenditures. Our central estimate is an income elasticity of 0.7, with an elasticity of 1.1 as the upper end of the 95 percent confidence interval. Point estimates from alternative specifications fall on both sides of our central estimate, but are almost always less than 1. We also present evidence suggesting that there are unlikely to be substantial national or global general equilibrium effects of rising income on health spending, for example through induced innovation. Our overall reading of the evidence is that rising income is unlikely to be a major driver of the rising health share of GDP.

The third chapter, written jointly with Kory Kroft, studies theoretically and empirically how optimal Unemployment Insurance (UI) benefits vary with local labor market conditions. Theoretically, we derive the relationship between the moral hazard cost of UI and the unemployment rate in a standard search model. The model motivates our empirical strategy which tests whether the effect of UI benefits on unemployment durations varies with the local unemployment rate. In our preferred specification, a one standard deviation increase in the local unemployment rate reduces the magnitude of the duration elasticity by 32%. Using this estimate to calibrate the optimal level of UI benefits, we find that a one standard deviation increase in the unemployment rate leads to a 6.4 percentage point increase in the optimal replacement rate.

JEL classification: J61, I10, J65.

Thesis Supervisor: David Autor
Title: Professor of Economics

Thesis Supervisor: Amy Finkelstein
Title: Professor of Economics

Acknowledgments

I have benefited greatly from the advice and support of my advisors Daron Acemoglu, David Autor, and Amy Finkelstein. I thank them for helping me make the transition from graduate student to researcher, and I am also thankful for their unwaveringly honest feedback throughout my time as a graduate student. I learned from them that the best research requires a great question, a creative answer, and a lot of extremely hard work.

I also want to thank my MIT classmates and other MIT faculty for giving helpful feedback at various stages of my research. I acknowledge many of them individually in each of the following chapters.

My interest in economics started in high school with the AP Economics course taught by Betsy Sidor. I thank her for stimulating my interest in economics as a field, and for challenging me to "think like an economist."

Lastly, I want to thank my family for their love and support. My wife Jenny was eternally supportive during the stressful search for a thesis topic, and she was (and is!) always willing to discuss my latest research ideas. My parents have always supported my academic studies, and I have tried to follow their advice to make the most of my educational opportunities.

Contents

Chapter 1

The Incidence of Local Labor Demand Shocks[1]

When a city experiences an adverse labor demand shock, the share of the adult population with a college degree tends to decline, as the net out-migration rate of college-educated workers exceeds non-college workers (Glaeser and Gyourko, 2005). A standard explanation for this pattern is that barriers to mobility are greater for low-skill workers (Topel, 1986; Bound and Holzer, 2000).[2]

This paper proposes and tests an alternative explanation which focuses on why low-skill workers may be disproportionately compensated during adverse labor demand shocks, rather than why it may be disproportionately costly for them to out-migrate. This explanation has two components. First, as documented below, adverse

[1]I am grateful to Daron Acemoglu, David Autor, and Amy Finkelstein for their guidance and support. I also thank Leila Agha, Josh Angrist, Fernando Duarte, Tal Gross, Cynthia Kinnan, Jean-Paul L'Huillier, Amanda Pallais, Jim Poterba, Michael Powell, Nirupama Rao, Bill Wheaton and participants at the MIT Labor and Public Economics Seminar for helpful comments. I gratefully acknowledge the National Institute of Aging (NIA grant number T32-AG000186) and the MIT Shultz Fund for financial support.

[2]The existence of greater barriers to mobility for low-skill workers is consistent with a large empirical literature that has documented that low-skill wages are more responsive than high-skill wages to local labor market conditions. For example, Bound and Holzer (2000) find that the elasticity of wages with respect to local labor demand is about 60% higher for workers with no more than a high school education than for college-educated workers. Similarly, Topel (1986) finds that local labor demand shifts generate much smaller wage differentials among more educated workers. Topel writes "consistent with the greater geographic mobility of more educated workers, their wages are less sensitive to both current and future changes in relative employment."

shocks substantially reduce the cost of housing. This fact and the existing evidence that the expenditure share on housing declines with income imply that low-skill workers are disproportionately compensated by housing price declines.[3] Second, means-tested public assistance programs disproportionately compensate low-skill workers during adverse shocks. I document below that, not surprisingly, aggregate transfer program expenditures are highly responsive to local labor market conditions.

These two different types of explanations – one based on mobility costs and one based on compensating factors – are not incompatible; however, their relative importance ultimately determines the actual incidence of local labor demand shocks. If out-migration of workers is low primarily because of mobility costs, then the incidence of local labor demand shocks will be primarily borne by workers; additionally, to the extent that mobility costs are greater for low-skill workers, they may disproportionately bear the incidence of the adverse shock. Alternatively, if the incidence of adverse local labor demand shocks is primarily borne by immobile housing and social insurance programs, then low-skill workers will be disproportionately compensated and, consequently, less likely to out-migrate.

In this paper, I develop and estimate a spatial equilibrium model which captures how wages, population, housing prices, and transfer payments re-equilibrate following a shift in local labor demand. The model is based on the spatial equilibrium model in Roback (1982). Following Glaeser and Gyourko (2005), the model in this paper allows for a concave local housing supply curve, arising from the durability of the local housing stock.[4] While the Glaeser and Gyourko model assumes perfect mobility, I allow for heterogeneous mobility costs which limit spatial arbitrage, as in Topel (1986). Unlike the preceding models, I explicitly model local labor demand.

To give the basic intuition of the model, consider the following simplified version.

[3] Of course, if low-skill workers are homeowners and not renters, then there is a negative wealth effect in addition to the decline in the user cost of housing following a negative local labor demand shock. Consistent with much of the recent urban economics literature (e.g., Glaeser and Gyourko (2005) and Moretti (2009)), I assume in the model below that everyone is a renter.

[4] Throughout the paper I use the term "concave housing supply curve" to imply that positive housing demand shocks increase housing prices less than equal-sized negative shocks reduce housing prices. More formally, a concave housing supply curve implies that $\partial^2(\text{housing price})/\partial(\text{housing supply})^2 < 0$.

Workers in a city inelastically supply labor so that net migration fully determines local labor supply. Workers do not differ in productivity, and there are no transfer payments.[5] Firms are perfectly mobile so that labor demand is perfectly elastic. Homogeneous housing units are supplied by absentee landlords who live in other cities, and workers consume a fixed expenditure share of housing (s_h). The main conceptual experiment in the model is that a single city experiences a (positive or negative) labor demand shock while a large number of other cities remain unchanged. Figures 1 and 2 provide graphical representations of the different equilibrium responses of wages, population and housing prices for four scenarios, depending on whether housing supply is constant elasticity or asymmetric and whether workers are perfectly mobile or face mobility costs when out-migrating.

Figure 1 depicts the equilibrium response when the elasticity of supply of housing is constant.[6] The figure shows a positive shift in the labor demand curve which raises wages by Δ. This increase in wages causes in-migration, which bids up housing prices until the increase in housing costs exactly offsets the wage increase (thus restoring the equilibrium no-arbitrage condition for workers). If workers are perfectly mobile, then the figure shows that the effect of a negative shock ($-\Delta$) is *symmetric*; i.e., wages, housing prices, and population adjust by equal and opposite magnitudes (as shown by L_-^A in the figure). This symmetry comes from the log-linearity of the housing supply curve and the perfect mobility of workers. If, alternatively, workers face non-negligible mobility costs, then there will be less out-migration following a negative shock. With non-negligible mobility costs, the no-arbitrage condition is now that the marginal worker must be indifferent between staying and paying c to out-migrate. In this case, both the population and housing price responses are *asymmetric*: positive shocks increase population and housing prices more than negative shocks reduce them (see L_-^B in the figure).

In Figure 2, the housing supply elasticity is no longer constant. Specifically, housing is more elastically supplied following an increase in housing demand than

[5]The full model below introduces high-skill and low-skill workers as well as transfer payments
[6]This is equivalent to assuming that the housing supply curve is log-linear.

a decrease in demand. As discussed in greater detail in the main text below and in the Appendix, this asymmetric housing supply curve is consistent with a simple model of durable housing where housing units are not destroyed once created (Glaeser and Gyourko, 2005). When workers are perfectly mobile, housing prices respond symmetrically (despite the asymmetry in the housing supply curve). Intuitively, housing costs still must adjust to exactly offset the wage changes. Only population responds asymmetrically (as shown by L_{-}^{C} in the figure). However, if workers have heterogeneous mobility costs to out-migrate as described above, then in this case the asymmetry of the population response is even greater (see L_{-}^{D} in the figure), and housing prices also respond asymmetrically.

These scenarios give the intuition for the following two implications of the model: (1) if positive labor demand shocks increase population more than negative shocks reduce population, this suggests the existence of a concave housing supply curve and/or heterogeneous mobility costs, and (2) if positive shocks increase housing prices more than negative shocks reduce housing prices, that is consistent with the existence of heterogeneous mobility costs.

The model guides the empirical strategy, which consists of two steps. In the first step, I test for asymmetric responses of wages, employment, population, and housing prices to symmetric labor demand shocks. The validity of this exercise requires constructing plausibly exogenous positive and negative shifts in local labor demand of equal magnitude. This paper follows Bartik (1991) in constructing an instrumental variable for local labor demand shocks by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. I find robust evidence using U.S. Census data that positive local labor demand shocks increase population (and employment) more than negative shocks reduce population (and employment) and that this asymmetry is greater for low-skill workers. These robust asymmetric relationships for local population and employment contrast sharply with the absence of any evidence of a similar asymmetric relationship for (any measure of) wages, housing values, and rental prices, though all of these other variables respond strongly to local labor demand.[7] As the spatial equilibrium model makes clear, these results are consistent with a concave local housing supply curve and limited mobility costs.

To quantitatively estimate the magnitude of mobility costs by skill and the shape of the housing supply curve, in the second set of empirical analyses I estimate the full model using a nonlinear, simultaneous equations GMM estimator. The GMM estimates suggest that the housing supply curve is concave and that (over decadal time horizons) mobility costs are not large and are comparable for both high-skill and low-skill workers.[8] The GMM results reveal several other important findings. First, the observed asymmetric population responses are primarily accounted for by an asymmetric housing supply curve rather than due to substantial barriers to mobility. Second, the results suggest that the observed difference in out-migration by skill is primarily accounted for by transfer payments rather than to differences by skill in housing expenditure shares. Lastly, the results suggest that the primary explanation for the comparative immobility of low-skill workers is not higher mobility costs, but rather a lower incidence of adverse local labor demand shocks. Consequently, much of the incidence of adverse labor demand shocks is diffused to homeowners, landlords, and public assistance programs.[9]

[7]The model in Glaeser and Gyourko (2005) predicts a concave relationship between housing prices and the exogenous labor demand, and these authors find supportive evidence of this prediction using an exogenous shock based on climate. As discussed in more detail in the Appendix, the key difference between the model in this paper and the model in Glaeser and Gyourko (2005) is that the model in this paper assumes that housing units are homogeneous, while in the Glaeser and Gyourko model housing units have heterogeneous, location-specific amenities. In other words, in the Glaeser and Gyourko model, exogenous shocks induce compositional changes in the distribution of location-specific amenities in the housing stock, and these compositional changes affect the (unconditional) average housing price. The difference in empirical results comes from the fact that Glaeser and Gyourko (2005) use mean temperature to construct local amenity shocks based on a dummy variable for whether or not the January mean temperature is greater than 29.1 degrees whereas I use variation in local labor demand.

[8]As discussed in more detail below, mobility costs are defined as a fraction of income, so that finding comparable mobility costs for high-skill and low-skill workers implies lower absolute mobility costs for low-skill workers.

[9]There is a related literature on the effect of income on migration (Kennan and Walker, 2009) and the effect of welfare decisions on the individual migration decision (Kennan and Walker, 2008). Both of these papers are highly complementary to this paper, as they employ a very different empirical approach. Kennan and Walker (2008) use NLSY data to estimate a rich structural model of migration. Their data set of welfare-eligible women with a high-school education contains 88 moves (out of 3,466 person-year observations), and the data are used to identify the effect of income

The estimation of the full model necessarily requires stronger assumptions than were needed to test for asymmetric responses to shocks. In order to be able to consistently estimate the *relative* magnitude of mobility costs by skill, I must assume that unobserved changes in local amenities induced by local labor demand shocks are not differentially valued by high-skill and low-skill workers. To be able to consistently estimate the *absolute* magnitude of mobility costs, however, a stronger assumption is needed; namely, that unobserved changes in local amenities are uncorrelated with local labor demand shocks. Because of this, the analysis of the absolute magnitudes of mobility costs should be treated as more speculative.

The rest of the paper proceeds as follows. Section 2 presents the theoretical framework. Section 3 discusses the empirical strategy and the data. Section 4 presents the reduced form empirical results. Section 5 investigates the robustness of these results. Section 6 presents GMM estimates of the full model. Section 7 concludes.

1.1 Theoretical Framework

This section presents a simple spatial equilibrium model of a local labor market that captures how wages, population, housing prices and transfer payments re-equilibrate following a labor demand shock.[10] The heart of the model is a no-arbitrage condition in which the marginal worker is indifferent between remaining in the city receiving the shock and moving away (Roback, 1982). This condition implicitly defines a local labor supply curve which determines the amount of migration in response to a labor demand shock. The model below allows for mobility costs, which limit spatial arbitrage and cause the incidence of the labor demand shock to at least partially fall on workers (Topel, 1986).[11] Additionally, the model admits two types of workers

on migration probability. Also related to this paper is the recent literature on the causal effect of education and geographic mobility (Wozniak, 2006; Malamud and Wozniak, 2008).

[10]The model is a "local general equilibrium" model in the sense that labor demand shocks affect non-labor markets within the city; however, it is not a full general equilibrium model because when the single city is shocked, the (minimal) effects on the rest of the universe are ignored.

[11]Topel (1986) is primarily concerned with understanding differences between permanent and transitory shocks; in the simple two-period model in this paper, all shocks are necessarily permanent.

(high-skill and low-skill) who differ in productivity, imperfectly substitute in production, and may also differ in their housing expenditure share, eligibility for transfer payments, and mobility costs. If an adverse labor demand shock causes relatively greater out-migration of high-skill labor, the model clarifies when this is because the incidence of the shock is borne by other factors that disproportionately compensate low-skill workers and when this is due to greater barriers to mobility for low-skill workers.

For simplicity, the model is presented as a two-period model in order to rule out the effects of long-run expectations, the differences between temporary and permanent shocks, option value from moving, and other issues arising in dynamic spatial equilibrium models. Between the two periods, a single city (out of a large universe of cities) experiences a labor demand shock between the first and second period.

To give the general intuition of the model, consider an adverse local labor demand shock in a city. This shock will reduce wages, which encourages out-migration and, ultimately, lowers housing prices until the no-arbitrage condition is restored for the marginal worker. The amount of out-migration is determined by the magnitude of mobility costs, the generosity of transfer payments, and the elasticity of supply of housing in response to a decline in housing demand.

The four main components of the model (labor demand, transfer payments, housing market, and labor supply) are now discussed in detail.

1.1.1 Labor Demand

Assume a large number of cities indexed by i, and define the (large) number of high-skill and low-skill workers in city i and time t as H_{it} and L_{it}. Production of the homogeneous tradable good y is given by the following CES aggregate production function:[12]

$$y_{it} = \theta_{it}((1-\lambda)L_{it}^{\rho} + \lambda(\zeta H_{it})^{\rho})^{\alpha/\rho}$$

[12]For simplicity, capital is not included in the model. This could be important if part of the incidence of labor demand shocks falls on renters of capital. Since the empirical results are based on decadal changes, it seems reasonable to assume that the elasticity of supply of capital over this time period is fairly large.

where λ is a share parameter, α measures the returns to scale of the labor aggregate, ζ is the relative efficiency of high-skill labor and ρ is related to the elasticity of substitution between high-skill and low-skill labor by $\sigma_{H,L} \equiv 1/(1-\rho)$.[13] The θ_{it} term is a city-specific index of local labor demand. In the empirical section below, I argue that my instrumental variable for local labor demand is a valid exogenous source of variation in θ_{it}.

Assuming wages are set on the demand curve, then they are given by the following marginal productivity conditions:

$$
\begin{aligned}
w_{it}^{H} &= \alpha\theta_{it}((1-\lambda)L_{it}^{\rho} + \lambda(\zeta H_{it})^{\rho})^{(\alpha-\rho)/\rho}\lambda\zeta(\zeta H_{it})^{\rho-1} \\
w_{it}^{L} &= \alpha\theta_{it}((1-\lambda)L_{it}^{\rho} + \lambda(\zeta H_{it})^{\rho})^{(\alpha-\rho)/\rho}(1-\lambda)(L_{it})^{\rho-1}
\end{aligned}
$$

Totally differentiating the above wage expressions results in the following conditions for the evolution of wages in terms of exogenous labor demand shock ($\Delta\theta_{it}$) and the endogenous migration responses (ΔH_{it} and ΔL_{it}):

$$
\Delta w_{it}^{H} = \Delta\theta_{it} + ((\rho-1) + (\alpha-\rho)(\pi))\,\Delta H_{it} + (\alpha-\rho)(1-\pi)\Delta L_{it} \tag{1.1}
$$

$$
\Delta w_{it}^{L} = \Delta\theta_{it} + ((\rho-1) + (\alpha-\rho)(1-\pi))\,\Delta L_{it} + (\alpha-\rho)(\pi)\Delta H_{it} \tag{1.2}
$$

where $\pi = \lambda(\zeta H)^{\rho}/((1-\lambda)L^{\rho}+\lambda(\zeta H)^{\rho})$, and the Δ operator represents the percentage change over time.

[13] Let μ be the share of high-skill workers in the labor market. Then if $\lambda = (1-\mu)^{\rho-1}/((\zeta\mu)^{\rho-1} + (1-\mu)^{\rho-1})$, ζ will give the equilibrium wage premium.

1.1.2 Transfer Payments

Means-tested public assistance programs are available only to low-skill workers and are modeled as a constant elasticity function of wages:[14]

$$t_{it} = \bar{T}(w_{it}^L)^{\Psi}$$

where t_{it} is the transfer income for the representative low-skill worker, $\bar{T}$ is a constant, and Ψ is the elasticity of public assistance income with respect to low-skill wages. The constant elasticity assumption is a simplification; empirically, I find no evidence of a nonlinear or asymmetric effect of labor demand shocks on transfer payment take-up, so this assumption appears to be reasonable. The equations above imply the following expression for the evolution of transfer income in response to changes in low-skill wages:

$$\Delta t_{it} = \Psi \Delta w_{it}^L \tag{1.3}$$

I assume $\Psi < 0$, which implies that transfer programs provide wage insurance. Define s_t^L as the share of total income that comes from transfer programs for low-skill workers; for high-skill workers, $s_t^H = 0$.

1.1.3 Housing Market

A homogeneous housing stock is supplied by absentee landlords, and the aggregate housing supply curve is given by $H^S(p_{it}^h)$, where p_{it}^h is the price of housing. Workers have identical non-homothetic preferences over housing and the homogeneous tradable consumption good. Most empirical estimates find that housing consumption is a normal good with an income elasticity of demand less than one. For example, Polinsky and Ellwood (1979), find a (permanent) income elasticity of $0.80 - 0.87$. These results suggest that the expenditure share of housing should be lower for high-

[14] Using PSID data from 1990, I calculate that 0.5% of households receiving AFDC income during the past year had a household head with at least a college degree. Among households receiving food stamps during the past year, the fraction is 0.7%. The percentages for a household head with a high school education or less are 79.1% (AFDC) and 82.6% (Food Stamps).

skill workers. Using data from the Consumer Expenditure Survey, this fact is clearly present in the cross-section: in 1995, the housing expenditure share declines by more than 8 percentage points going from bottom 20% in income to top 20% in income distribution, declining from 38.5% to 30.0%.[15] Defining s_h^H and s_h^L as the housing expenditure shares for high-skill and low-skill workers, respectively, then these facts indicate that $s_h^L > s_h^H$.

Rather than specifying a specific functional form to derive an expression for aggregate housing demand, I instead approximate housing demand as follows:

$$H^D(p_{it}^h) = \frac{s_h^H w_{it}^H H_{it} + s_h^L(s_t^L t_{it}^L + (1 - s_t^L)w_{it}^L)L_{it}}{p_{it}^h}$$

This expression is an approximation since I am implicitly assuming that any changes in income induced by a shift in labor demand are small so that income effects can be ignored. Empirically, the changes in wages within skill groups are small relative to the differences in wages across skill groups.

The initial supply-demand equilibrium in housing market in the first period is given by $H^S(p_{it}^h) = H^D(p_{it}^h)$. Totally differentiating this equilibrium condition gives the following expression for the housing market response:

$$\Delta p_{it}^h + \Delta H^S(\Delta p_{it}^h) = \nu(\Delta y_{it}^H + \Delta H_{it}) + (1 - \nu)(\Delta y_{it}^L + \Delta L_{it}) \tag{1.4}$$

where ν is the high-skill share of aggregate housing demand and Δy_{it}^j gives the change in total income for skill group j ($\in \{H, L\}$); i.e., $\Delta y_{it}^j = s_t^j \Delta t_{it}^j + (1 - s_t^j)\Delta w_{it}^j$. If the housing supply curve has constant elasticity, then $\Delta H^S(p_{it}^h) = \sigma \cdot \Delta p_{it}^h$. Since housing is a durable good, however, the housing supply elasticity is not likely to be constant. Instead, the housing supply elasticity will be larger for increases in housing demand than for decreases in housing demand due to the durability of the housing stock (Glaeser and Gyourko, 2005). Formally, durable housing implies that $\Delta H^S(\Delta p_{it}^h)$ is increasing in Δp_{it}^h. The Appendix presents a simple model which pro-

[15]Expenditure share by quintile (going from lowest to highest income quintile) is the following: 38.5%, 32.9%, 31.8%, 30.0%, and 30.0%.

vides microfoundations for a concave housing supply curve based on slow depreciation of the housing stock and a heterogeneous distribution of costs of supplying housing.

1.1.4 Labor Supply

For simplicity, I assume that workers inelastically supply labor to their local labor market, so that all variation in local employment comes only from migration decisions. The local labor supply curve is then implicitly defined by a mobility condition which states that the marginal migrant must be indifferent between remaining in city i and moving to any other city.

I introduce costly spatial arbitrage by assuming that workers have heterogeneous mobility costs. I construe mobility costs broadly to encompass both financial and psychic barriers to out-migration as well as heterogeneous tastes and distastes for a given location. Thus unlike Topel (1986), I allow mobility costs to take on positive and negative values. Positive values encompass both actual moving costs as well as preferences for the current city, while negative values represent distaste of potential in-migrants for a given area. Formally, I model this by assuming that mobility costs for workers in city i are independently drawn from distributions $M_i^H(m)$ and $M_i^L(m)$ (with support $[0, \infty)$), while the mobility costs of in-migrating into city i for the workers living in all of the other cities are drawn from the distributions $M_{-i}^H(m)$ and $M_{-i}^L(m)$ (with support $(-\infty, 0]$).

These mobility cost distributions imply mobility cost functions $c^H(\Delta H_{it})$ and $c^L(\Delta L_{it})$, which return the mobility cost of the marginal migrant given the change in population between the first and second period. Mobility costs are defined as a fraction of total income, so that the marginal migrant receiving $(w + t)$ in city i will pay $(w + t)c$ to out-migrate. For a smooth distribution of mobility costs, the mobility cost function will be strictly decreasing, so that the mobility cost of the marginal migrant increases as more workers out-migrate.[16]

[16] Note that this two-period model contains two important simplifications which make it straightforward to study mobility costs. First, following Topel (1986), gross migration will always equal net migration, so that there is only one marginal migrant per worker type in each city. The work of Artuc, Chaudhari, and McLauren (2009) and Chaudhari and McLauren (2007) suggest a tractable

To derive the (implicit) labor supply curve for low-skill workers, let $v_i(w_{it}^L + t_{it}^L, p_{it}^h)$ be the indirect utility function for the marginal low-skill worker in city i. Spatial equilibrium in the first period requires that the following condition holds for the marginal low-skill migrant in city i:

$$v_i(w_{it}^L + t_{it}^L, p_{it}^h) = v_j(w_{jt}^L + t_{jt}^L, p_{jt}^h) \; \forall j \neq i$$

Now consider a shock to θ_i in city i. The shock will cause a wage differential which will encourage costly migration to arbitrage the wage and employment differential, and the price of housing and transfer payments will also adjust as a local general equilibrium response to the shock. Differentiating the above spatial equilibrium condition and applying Roy's Identity results in the following expression:

$$(1 - s_t^L)\Delta w_{it}^L + s_t^L \Delta t_{it}^L - s_h^L \Delta p_{it}^h + c^L(\Delta L_{it}) = 0 \tag{1.5}$$

where s_t^L $(= t^L/(w^L + t^L))$ is public assistance income as a share of total income. An analogous expression holds for high-income workers (where $s_t^H = 0$):

$$\Delta w_{it}^H - s_h^H \Delta p_{it}^h + c^H(\Delta H_{it}) = 0 \tag{1.6}$$

Equations (1.5) and (1.6) are implicit labor supply curves because net migration is determined by the spatial equilibrium condition for the marginal migrant. In words, the conditions above state that the change in indirect utility in response to changes in wages, transfer payments, and housing prices must equal the mobility costs of the marginal migrants. The ΔL_{it} and ΔH_{it} terms represent the amount of net migration that needs to occur to make these two equations hold.

These two equations highlight the three reasons discussed in the introduction why net migration rates may differ by skill. First, public assistance programs are

way to relax this assumption and allow gross migration flows to exceed net migration flows. Second, the mobility cost function is allowed to be asymmetric, but since this is a two-period model the shape of this function does not depend on the history of past shocks. In a fully dynamic model, the history of past shocks may affect the elasticity of supply of in-migrants and out-migrants.

means-tested, so that $s_t^L > s_t^H = 0$. Second, as documented below, low-skill workers consume a larger fraction of their income on housing $s_h^L > s_h^H$, meaning that housing price declines disproportionately compensate low-skill workers. Finally, the mobility cost functions may differ by skill. If low-skill workers typically face higher mobility costs following a negative shock, then $c^L(x) > c^H(x)\ \forall x < 0$.

1.1.5 Equilibrium

Following an exogenous shock to local labor demand ($\Delta\theta_{it}$), the new equilibrium of the model is defined by the following conditions:

- Labor demand adjusts so that high-skill and low-skill wages equal marginal products (equations (1.1) and (1.2))
- Transfer payments adjust according to changes in low-skill wages (equation (1.3))
- Housing prices adjust so that the change in housing demand equals the change in housing supply (equation (1.4))
- Population adjusts so that the marginal high-skill and low-skill migrant is indifferent between staying and leaving (equations (1.5) and (1.6))

Although the nonlinearities in the housing supply curve ($\Delta H^S(\Delta p_{it}^h)$) and the mobility cost functions ($c^H(\Delta H_{it})$ and $c^L(\Delta L_{it})$) preclude analytical solutions without particular functional form assumptions, the Appendix derives comparative statics for specific scenarios under the special case of constant returns to scale of production ($\alpha = 1$).

Figure 3 reports results from simulating the model.[17] The figure shows that if population responds asymmetrically, it suggests the existence of a concave housing supply curve and/or the existence of heterogeneous mobility costs. The responsiveness of housing prices isolates the importance of heterogeneous mobility costs, since

[17] The details of the simulation are given in the Appendix.

mobility costs cause immobile workers to bid up the price of housing during negative shocks, causing housing prices to respond asymmetrically. Therefore, the model suggests that it is possible to identify both mobility costs and the shape of the housing supply curve by using information on the joint responses of wages, population, housing prices, and transfer payments to exogenous labor demand shocks. Empirically, I will first estimate nonlinear reduced form regressions to test for asymmetric responses to labor demand shocks, and I next carry out a full estimation of the model to recover the parameters which govern the distribution of mobility costs and the shape of the housing supply curve.

1.2 Empirical Strategy and Data

As the model makes clear, the reduced form relationships between each of the endogenous variables (Δw^H, Δw^L, ΔH, ΔL, Δp^h, Δt^L) and the labor demand shock $\Delta\theta$ are informative about the shape of housing supply curve and the presence of heterogeneous mobility costs. This motivates the following reduced form estimating equation:

$$\Delta x_{it} = g^x(\Delta\theta_{it}) + \alpha_t + \nu_{it}$$

where i indexes cities, t indexes time periods, x is one of the endogenous variables above, α_t captures proportional shocks to all cities in a given time period, ν_{it} is an error term, and $g()$ is a function to be estimated. Nonparametric estimates of $g()$ are reported graphically below. In addition to the nonparametric estimates, I also parameterize $g^x(\Delta\theta)$ as $\beta(\Delta\theta)+\delta(\Delta\theta)^2$ which leads to the following baseline reduced form empirical specification that is reported in the tables:

$$\Delta x_{it} = \beta \times \Delta\theta_{it} + \delta \times (\Delta\theta_{it})^2 + \alpha_t + \nu_{it} \tag{1.7}$$

where x is the endogenous variable of interest, β and δ are the coefficients on a quadratic in $\Delta\theta_{it}$, and α_t are year fixed effects. This reduced form specification is estimated by OLS using a proxy for local labor demand (described below). The quadratic specification allows the elasticity of x_{it} with respect to θ_{it} to vary: specifically, the elasticity at $\Delta\theta_{i,t} = 0$ is given by $\hat{\beta}$, while $\hat{\beta} + 2\hat{\delta}\Delta\theta_{it}$ is the elasticity at $\Delta\theta_{it}$. Since the equation is estimated in first differences it implicitly controls for time-invariant differences across geographic areas, while the inclusion of year fixed effects captures any (proportional) changes in x_{it} common to all cities. Formally, the statistical test of $\delta \neq 0$ is sufficient to establish that positive and negative shifts in labor demand of equal magnitude have unequal effects. However, this test is evaluating the null hypothesis of a linear relationship against a specific parametric alternative. Therefore, I will also report nonparametric specification tests which test the null hypothesis of a linear relationship against a nonparametric alternative (Ellison and Ellison, 2000).[18]

Lastly, I also estimate the full model developed above to recover flexible estimates of the mobility cost functions of high-skill and low-skill workers and the housing supply curve parameters. The estimation is a nonlinear, simultaneous equations problem, and it is implemented using a two-step optimal GMM estimator. The details of this procedure are described in more detail below.

1.2.1 An Omnibus Instrumental Variable for Local Labor Demand

In order to estimate equation (1.7) above, a valid instrumental variable for local labor demand is needed. I follow the empirical strategy of Bartik (1991) and construct a measure of plausibly exogenous labor demand shocks derived by interacting cross-sectional differences in industrial composition with national changes in industry employment shares.[19] This relative demand index can be used to predict changes in wages and employment. The identifying assumption is that changes in industry shares at the national level are uncorrelated with city-level labor supply shocks and

[18]I view these tests as complementary to the significance tests of the quadratic term; while the nonparametric specification tests do not require formulating a specific parametric alternative, it is difficult to ensure that these tests have the right size and power.

[19]See Blanchard and Katz (1992), Bound and Holzer (2000), Autor and Duggan (2002), and Luttmer (2005) for other applications of this instrumental variable.

therefore represent plausibly exogenous (demand-induced) variation in metropolitan area employment. This predicted employment variable ($\hat{E}_{it}$) is used to create a predicted change in local area employment ($\Delta\hat{\theta}_{it}$) as follows: $\Delta\hat{\theta}_{i,t} = (\hat{E}_{it} - E_{i,t-\tau})/E_{i,t-\tau}$. This measure is used as a proxy for $\Delta\theta_{it}$.[20]

The key identifying assumption is that this proxy is uncorrelated with unobserved shocks to local labor supply. In this paper a stronger assumption is also needed – specifically, I must assume that $\Delta\theta_{i,t} = \Delta Z$ and $\Delta\theta_{i,t} = -\Delta Z$ represent shifts in local labor demand of equal magnitude. This requirement gives one advantage of the Bartik procedure over other identifiable shocks to local labor demand, since this instrumental variable is an *omnibus* measure of changes in local labor demand. By contrast, if one were to use identifiable shifts to labor demand such as movements in oil prices, coal prices, or other natural resource shocks it would require that equal-sized positive and negative price changes represent equal-sized shifts in local labor demand. This may be difficult to justify in natural resource industries that are typically characterized by high amounts of specific capital and/or irreversible investments. A related benefit of the Bartik procedure is that subsets of industries can be excluded when constructing the instrumental variable to verify that the results are not driven by particular sectors.

1.2.2 Data and Descriptive Statistics

The data sources are briefly described here. The Appendix gives more detail on how the data set was created.

Census Integrated Public Use Microsamples (IPUMS) The basic panel of metropolitan area data comes from the 1980, 1990, and 2000 Census individual-level

[20]Formally, predicted employment growth is computed as follows:

$$\begin{aligned}\pi_{it} &= \sum_{k=1}^{K} \varphi_{i,k,t-\tau} \left(\frac{\upsilon_{-i,k,t} - \upsilon_{-i,k,t-\tau}}{\upsilon_{-i,k,t-\tau}} \right) \\ \hat{E}_{it} &= (1 + \pi_{i,t}) n_{i,t} \\ \Delta\hat{\theta}_{it} &= (\hat{E}_{it} - E_{i,t-\tau})/E_{i,t-\tau}\end{aligned}$$

where $\varphi_{i,k,t-\tau}$ is the employment share of industry k in city i and $\upsilon_{-i,k,t}$ is the national employment share of industry k excluding city i.

and household-level extracts from the IPUMS database (Ruggles et al, 2004).[21] The baseline data are limited to individuals and households living in metropolitan areas. The IPUMS data are used to construct estimates of local area wages, employment, population, housing prices, and rental prices in each metropolitan area. The primary advantage of the Census data is the ability to construct city-level measures disaggregated by skill. These data are also used to construct the predicted labor demand instrumental variable by using the industry categories of the individuals in the labor force. See the Appendix for remaining details.

Regional Economic Information System (REIS) The metropolitan-area measures of expenditures on public assistance programs are computed by aggregating the county-level aggregate data in the REIS. The REIS contains annual county-level data on total expenditures broken down by transfer program (e.g., food stamps, income maintenance programs, public medical benefits, veterans benefits, SSI benefits). Counties are aggregated into metropolitan areas using the 1990 Metropolitan Statistical Area (MSA) definitions. Because of the difficulty in aggregating counties into MSAs within Alaska and Virginia during this time period, MSAs in these states are dropped from the baseline sample. Though the data are not disaggregated below the county-level, the data are based on government agency reports and are therefore quite reliable. Additionally, according to recent work by Meyer, Mok, and Sullivan (2009), aggregate expenditure data may be sometimes preferable to individual or household survey data due to substantial underreporting in the latter. All transfer program measures are adjusted per low-skill capita based on the non-college adult population.[22]

Table 1 reports descriptive statistics for the final data set.

[21]The 2007 American Community Survey (ACS) is included as a robustness check. The 1970 Census is not used at all because it identifies only a small subset of the MSAs that appear in later years.

[22]Meyer, Mok, and Sullivan (2009) find substantial underreporting of benefit receipt in a wide range of data sets, including the CPS, PSID, SIPP, PSID, and the Consumer Expenditure Survey for a wide range of transfer programs. They also document that the under-reporting is not consistent over time.

1.3 Results

1.3.1 Graphical Evidence

Figures 4 and 5 report nonparametric reduced form estimates for the primary dependent variables. In addition to the nonparametric estimates, linear estimates are graphed for comparison. The figures also display bootstrapped (uniform) 95% confidence intervals.[23] The confidence intervals are very wide at the extremes, which makes it difficult to reject the null hypothesis that the data are described by a linear relationship. However, in some cases the confidence intervals reject the specific linear relationship estimated using a parametric linear model, though this visual test ignores estimation error in the linear model. Consequently, the nonparametric specification tests reported below will be useful in assessing whether the data reject the null hypothesis that the parametric linear model is appropriate.[24]

Overall, across all of the graphs the only suggestive evidence of an asymmetric response is for employment, population, and transfer payments. The population and employment graphs show a convex relationship with the labor demand instrumental variable. By contrast, there is no evidence of a similar asymmetric relationship for housing values, rental prices, or any measure of wages (wage measures are defined below). As shown by the simulated data in Figure 3, these results are consistent with a concave housing supply curve and limited mobility costs. In order to formally test for the existence of an asymmetric response (and measure the magnitude of the asymmetry when it exists), the next subsection reports results from quadratic specifications and nonparametric specification tests.

[23] The bootstrapped confidence intervals are computed based on 10,000 replications, where MSAs are sampled with replacement. In each bootstrap step, an undersmoothed local linear bandwidth is chosen following Hall (1992). That paper reports Monte Carlo results which suggest that undersmoothing produces confidence interval estimates with greater coverage accuracy than confidence intervals obtained by explicit bias correction. The bandwidth of the Epanechnikov kernel used for point estimation is 0.041; the undersmoothed kernel bandwidth is $0.75 \cdot 0.041 = 0.031$.

[24] In all figures, the nonparametric estimates are local linear regressions. The nonparametric reduced form estimates are also constrained to be monotonic following the rearrangement procedure of Chernozhukov, Fernandez-Val, and Galichon (2003). The rearranged estimates are more efficient under the null hypothesis that the true relationship is (weakly) monotonic. In all figures, the unconstrained estimates are qualitatively similar.

1.3.2 Reduced Form Results

This section reports estimates of equation (1.7) above to investigate the responsiveness of wages, employment, and population to changes in local labor demand. The baseline reduced form estimating equation is reproduced below:

$$\Delta x_{it} = \beta \times \Delta\hat{\theta}_{it} + \delta \times (\Delta\hat{\theta}_{it})^2 + \alpha_t + \Delta\nu_{i,t}$$

The results are reported in Tables 2 through 4. Table 2 presents results for overall population, employment, and wages. Column (1) shows the results for the total population between the ages of 18 and 64.[25] The estimate of β is precise and strongly statistically significant ($p < 0.001$), which verifies that the measure of predicted employment changes strongly predicts actual shifts in local population. The estimate of δ is also precise and strongly statistically significant. The estimate is positive and large in magnitude ($\hat{\delta} = 28.004$). One way to interpret the magnitude of this estimate is to calculate the marginal effect at one standard deviation greater than zero and one standard deviation less than zero; these estimates are -0.115 and 3.716, respectively, and the difference between these estimates is strongly statistically significant ($p < 0.001$).[26] Additionally, a nonparametric specification test strongly rejects the null hypothesis that the relationship is linear in favor of a nonparametric alternative ($p < 0.001$).[27] In other words, the results in this column suggest that positive changes in local labor demand increase population more than negative changes reduce population. The results for employment in column (2) show evidence of a similar convex relationship. The results in column (3) using the percentage point

[25] Results using the population between the ages of 25 and 54 are very similar.

[26] Note that the p-value on the test for whether the marginal effects are the same at one standard deviation above and below zero is the same as the p-value on the test of whether the quadratic term is statistically significantly different from zero.

[27] I use the nonparametric specification test procedure suggested by Ellison and Ellison (2000), which groups the data into "bins" and creates a test statistic that is asymptotically distributed as a standard normal random variable. To my knowledge, there is a not a data-driven procedure to select the proper bin width; therefore, I view the nonparametric specification test as complementary to the quadratic specification. While the nonparametric specification test does not rely on a specific parametric alternative, it is not possible to ensure that I have the right size and power in constructing my statistical tests. In almost all of the results that follow, inference based on the quadratic specification and the nonparametric specification test is similar.

change in the employment-to-population ratio show that not all of the reduction in local employment from an adverse shock comes from net out-migration; there is also a decline in labor force participation.

The remaining columns of Table 2 explore the consequences of local labor demand shifts on wages. There are two difficulties in finding an appropriate wage measure. The first difficulty is that the labor demand shock may induce compositional changes in the population, so that the change in the average wage will be confounded by compositional effects. The second difficulty is that changes in labor force participation reduce income per adult, but would be excluded using a measure of average wages based only on employed workers. I approach this problem by first presenting two measures of changes in wage income which I believe represent upper and lower bounds of the true change in income holding characteristics of the workers fixed. The first measure (following Bound and Holzer (2000)) is the total wage income per 18-64 adult. This measure will account for demand-induced changes in labor force participation but will also include compositional changes. The results are in column (4) and show a large effect of local labor demand on wages ($\hat{\beta} = 0.959$). The second measure uses the individual-level census data and regresses log wages of employed workers on a large set of controls and MSA fixed effects (see Appendix for details). The MSA fixed effect estimated from this regression is a composition-adjusted measure of the wage premium which I define as the "residualized wage".[28] The results in column (5) using this measure show a much smaller wage response ($\hat{\beta} = 0.353$). However, this second measure does not account for changes in labor force participation rate. Assuming that at least some of the observed change in labor force participation is involuntary, then this measure will understate the total effect. To address this concern, I take the residualized wage measure and multiply it by the observed labor force participation rate.[29] I call this the "adjusted wage" and use this as the preferred

[28]This measure is similar to the local wage premiums calculated in Shapiro (2003) and Albouy (2009a, 2009b). This measure does not control for unobservable changes in the composition of labor force. If unobservable changes in composition of labor force move in the same direction as observable changes, then the measured response of wages will be upward biased, and estimates of mobility costs will be conservative.

[29]Note that when I present results by skill below, I use the labor force participation rate in the given skill group to adjust the residualized wage measure.

wage measure. This measure accounts for both compositional changes in the labor force in response to the shock as well as changes in labor force participation. Note that if not all of the observed change in labor force participation is involuntary, I will be overstating the importance of mobility costs when I ultimately estimate the full model via GMM. Essentially, the adjusted wage measure assumes that reservation wages are negligible.[30]

As expected, the magnitude of the effect of local labor demand for adjusted wages lies in between the other two wage measures ($\hat{\beta} = 0.520$). Since the magnitude of changes in labor force participation is modest, the estimates for adjusted wages are closer to the estimates for residualized wages than the estimates using the per capita income measure. Regardless of the measure of wages used, however, the important conclusion that emerges from columns (4) through (6) is that there is no evidence of an asymmetric response of wages to shifts in local labor demand in any of the measures. It is only population and local employment which respond asymmetrically.

Table 3 reports results on population, employment and wages separately for high-skill and low-skill workers. I define low-skill workers as those without a college degree, and high-skill workers as those with at least a college degree. The patterns in Table 2 are reproduced when looking within each skill group: population and employment respond asymmetrically, and there is no evidence of a similar asymmetric response for either high-skill or low-skill wages. Additionally, columns (3) and (6) show that the skill composition of the adult population and labor force also responds asymmetrically. In other words, negative shocks reduce college share of adult population more than positive shocks increase college share.

Next, Table 4 looks at three important non-labor outcomes: housing values, rental prices, and aggregate expenditures on public assistance programs. The measures of average housing values and rental prices are purged of observable changes in the quality of the housing stock following a similar procedure to the one used to create

[30]As a way of bounding the estimated magnitude of mobility costs, I also report GMM estimates below which use the residualized wage instead of the adjusted wage. Under the assumption that reservation wages are less than offered wages for at least some adults, the residualized wage will give a lower bound on the estimated magnitude of mobility costs.

the residualized wage measure (see Appendix for details), though the results using the unconditional average housing values and rental prices are very similar. Column (1) reports results for housing values, which respond strongly to local labor demand. The results for rental prices are similar in magnitude and more precise. As with the wage results, there is no evidence of an asymmetric response. The estimates of δ in both columns (1) and (2) are statistically insignificant and at most modest in magnitude, and the nonparametric specification tests fail to reject the null hypothesis that the deviations from the parametric (linear) model are due to chance.

Column (3) reports estimates using aggregate expenditures on Food Stamps and Income Maintenance Programs. The results show that expenditures on these programs respond strongly to local labor market conditions. The estimated magnitude of the response is large ($\hat{\beta} = -2.366$) and implies that a 1% decline in local labor demand increases aggregate expenditures on these two programs by 2.4%.[31]

A setting in which population and employment respond asymmetrically to positive and negative labor demand shocks while wages, rental prices, and housing values respond symmetrically is consistent with the model simulation where mobility costs are limited and the housing supply curve is concave. Before moving beyond this qualitative conclusion to quantitative estimates of mobility costs and housing supply curve parameters, I next document that these reduced form results are not driven by unobserved trends, outliers, sample selection, or heterogeneous industry-specific effects. After that, I conclude by estimating the full model above using a nonlinear GMM estimator.

[31] Appendix Table A2 reports estimates for various other transfer programs, including Medicare, Disability Benefits, SSI, and Veterans Benefits, and the results are qualitatively similar. I focus on Food Stamps and Income Maintenance income because these programs are explicitly designed to smooth consumption.

1.4 Robustness

1.4.1 Industry Trends

The main results in Table 2 emphasize the importance of asymmetric employment and population responses to local labor demand shocks, and the absence of a similar asymmetric response for wages, housing prices, and transfer payments. The key identifying assumption in interpreting these results is that equally-sized positive and negative predicted changes in local employment represent shifts in local labor demand of plausibly equal magnitude. Because the predicted changes are formed by interacting cross-sectional variation in industrial composition with national changes in industry shares, an obvious concern is that qualitatively different industries are declining and expanding. If these industries would not be expected to have otherwise identical responses to shifts in local labor demand (perhaps because of differences in relative demand for high-skill labor, the amount of specific human capital in the industry, or the ability of firms in the industry to respond and adjust to shocks), then this would cast doubt on the interpretation of the results as tracing out an asymmetric local labor supply curve.

To investigate this concern, I categorize industries based on their decadal changes in total national employment. Industries are grouped into one of four categories:

1. *Persistently expanding industries.* Industries where employment increased every decade.

2. *Persistently declining industries.* Industries where employment decreased in every decade.

3. *Stable industries.* Industries where employment did not increase or decrease more than 20% in any of the decades.[32]

[32]If industries are classified as both persistently expanding/declining and stable, I categorize the industry as stable. This definition and the cutoff of 20% were chosen to give roughly equal-sized categories. Results are similar with nearby cutoffs.

4. *Volatile industries.* Industries that experienced employment growth of more than 20% *and* decreases of more than 20% during the sample period.

5. *Other industries.* Industries not otherwise categorized.

The top twenty industries according to average national employment share in each of these categories are listed in Table A1. The industries in each of the categories conform to expectations given the secular industry trends during this time period. Persistently expanding industries are concentrated in services, health care, data processing, and leisure goods, while persistently contracting industries are in apparel, publishing, manufacturing, and tobacco. Volatile industries include natural resource industries such as oil and gas extraction as well as defense industries. I begin by constructing predicted employment excluding variation in national employment shares for industries that are persistently expanding or persistently declining.[33] The resulting relative demand index is purged of any variation caused by secular trends in services and manufacturing. Table 5 reports results from estimating equation (1.7) using this alternative measure of predicted employment as an instrumental variable for local labor demand. The dependent variable is the change in adult population in all columns. Column (1) reproduces the results from column (1) in Table 2 for comparison. Column (2) reports results using the predicted employment measure that does not use any variation from industries which are persistently expanding or persistently declining. The point estimates in column (2) are fairly similar to the baseline estimates reproduced in column (1). Columns (3) through (5) report results excluding each of the other industry categories when constructing predicted employment, and the results are also quite similar to the baseline results in column (1). I interpret these results as suggesting that the estimated asymmetric population response is not primarily caused by heterogeneous industry-specific effects

[33] Formally, predicted employment growth is computed by using only the subset of industries which pass a given filter:

$$\pi'_{i,t} = \sum_{k \in K'} \varphi_{i,k,t-\tau} \left(\frac{v_{-i,k,t} - v_{-i,k,t-\tau}}{v_{-i,k,t-\tau}} \right)$$

where K' is the set of industries which pass the filter.

A related concern is that because of the way that the IPUMS creates consistent industry codes across time, there are "catch-all" industry codes that collect industries which are not otherwise categorized. I label an industry code a catch-all industry code if it contains the word "miscellaneous" or contains the suffix "not elsewhere categorized." These catch-all industry codes make up roughly 10% of the industry codes. These catch-all categories may represent different collections of industries in different decades, which may bias the main estimates. To investigate this concern, I create an alternative measure of predicted employment which does not use any variation in national employment shares of these industries. The estimates using this predicted employment measure are reported in column (6) and are similar to the results in column (1), suggesting that there is no significant bias from including these catch-all categories.

Tables 6 and 7 report results which repeat this exercise using as the dependent variable adjusted wages and rental prices, respectively. Consistent with the baseline results in Tables 2 and 4, none of the estimates in any of the columns show any evidence of an asymmetric relationship between adjusted wages or rental prices and labor demand.[34]

1.4.2 Alternative Specifications

I next turn to an investigation of the robustness of the main results by reporting alternative specifications which vary the sample definition and the set of time-varying controls used. The purpose of these specifications is primarily to investigate the possibility of sample selection bias and the potential bias from unobserved trends that are correlated with shifts in local labor demand. As with Tables 5 through 7, Tables 8 through 10 use population, adjusted wages, and rental prices (respectively) as the dependent variables. All columns report results from estimating variants of equation (5). In all tables, column (1) reports the baseline results for comparison.

[34]Interestingly, the magnitude of the (linear) response of adjusted wages and rental prices to local labor demand varies somewhat depending on the industries used to generate predicted changes in employment, suggesting that the strength of the proxy for local labor demand may vary depending on the set of industries used to generate the proxy.

Column (2) reports results from adding data on the 2000-2007 changes.[35] Column (3) creates "pseudo-MSAs" by grouping together all individuals in a state who are not in an MSA. Columns (4) and (5) report results including alternative sets of geographic and time fixed effects. Column (4) includes region fixed effects for each of the nine census regions which control for region-specific linear time trends. Column (5) includes controls for MSA-specific linear time trends. Column (6) reports results which test for the importance of outliers. This column drops the 5% of the data with the largest magnitude changes in local labor demand. Finally, column (7) uses the County Business Patterns (CBP) data set to construct the local labor demand instrument rather than using Census data (see Appendix for details). The CBP data contain finer industry categories, which in principle could reduce measurement error in the instrument, but there are two primary drawbacks: first, there is a high rate of suppressed data at the county-by-industry level, and, second, the county-level data must be aggregated. Dealing with both of these drawbacks introduces measurement error.

Table 8 reports results using population as the dependent variable. Across all of the columns, the point estimates are very similar to the baseline specification in column (1). The results in column (5) which include MSA-specific linear time trends show a substantial loss of precision, but the point estimates remain stable. The results in column (6) show that the estimated asymmetric response is robust to dropping outlying observations, suggesting that the convex population response is not primarily driven by outliers. The results in column (7) show the results are similar using CBP data to construct the labor demand instrument.

Tables 9 and 10 report results using adjusted wages and rental prices (respectively) as the dependent variables. The estimates of δ are never statistically significant at conventional levels, nor are even consistently the same sign across columns. In other words, there is no consistent evidence of an asymmetric response of adjusted wages or rental prices to local labor demand shocks.

[35]The 2000-2007 changes are translated into implied decadal changes by first calculating annual percentage changes.

Lastly, Appendix Table A3 reports specifications which drop each one (of nine) census regions. This table confirms that the results do not appear to be driven by any particular region.

In summary, the reduced form patterns of a significant asymmetric response of population and employment to changes in local labor demand appear robust and contrast sharply with a lack of similar asymmetric responses for wages, housing values, and rental prices.

1.5 GMM Estimates

The reduced form results presented above directly test for the existence of asymmetric responses of wages, population, employment, and housing prices to symmetric labor demand shocks. These results do not directly estimate any of the economic parameters in the theoretical model and are therefore not precisely informative about the distribution of mobility costs by skill and the actual incidence of labor demand shocks. This section reports results from a joint estimation of the full model using a nonlinear, simultaneous equations GMM estimator. The econometric setup follows from the theoretical model presented above and imposes moment conditions which can be used to identify the economic parameters of interest. In particular, the GMM estimator can recover flexible estimates of the housing supply curve and mobility cost functions for high-skill and low-skill workers. These estimates can be used to assess the relative importance of housing expenditures, transfer payments, and mobility costs in generating the observed migration patterns in the data. Additionally, because I parameterize the model so that there are more moment conditions than (remaining) parameters to estimate, the GMM estimator admits a chi-squared overidentification test of the full model.

To implement the GMM estimator, the following equations (derived from equations (1) through (6) in the model above) are used:

$$\Delta e_{it}^{wH} = \Delta w_{it}^{H} - (\Delta\theta_{it} + ((\rho-1) + (\alpha-\rho)(\pi))\,\Delta H_{it} + (\alpha-\rho)(1-\pi)\Delta L_{it})$$
$$\Delta e_{it}^{wL} = \Delta w_{it}^{L} - (\Delta\theta_{it} + ((\rho-1) + (\alpha-\rho)(1-\pi))\Delta L_{it} + (\alpha-\rho)(\pi)\Delta H_{it})$$
$$\Delta e_{it}^{t} = \Delta t_{it}^{L} - \Psi \Delta w_{it}^{L}$$
$$\Delta e_{it}^{h} = \Delta p_{it}^{h} + \Delta H^{s}(\Delta p_{it}^{h}) - (\nu(\Delta w_{it}^{H} + \Delta H_{it}) + (1-\nu)((1-s_t^L)\Delta w_{it}^{L} + s_t^L \Delta t_{it}^{L} + \Delta L_{it}))$$
$$\Delta e_{it}^{H} = \Delta w_{it}^{H} - s_h^H \Delta p_{it}^{h} + c^H(\Delta H_{it})$$
$$\Delta e_{it}^{L} = (1-s_t^L)\Delta w_{it}^{L} + s_t^L \Delta t_{it}^{L} - s_h^L \Delta p_{it}^{h} + c^L(\Delta L_{it})$$

where i indexes cities, t indexes time, and Δe_{it}^{j} represent error terms uncorrelated with shifts in labor demand.[36] These equations jointly solve the local general equilibrium problem of how wages, employment, housing prices, and transfer payments respond to an exogenous labor demand shift $\Delta\theta_{it}$. The six endogenous variables are the following: Δp_{it}^{h}, Δw_{it}^{H}, Δw_{it}^{L}, ΔH_{it}, ΔL_{it}, and Δt_{it}^{L}. Note that the error terms are allowed to be freely correlated with each other, which can give rise to simultaneity bias that the GMM estimator is intended to address. The unknowns in the model are the following parameters and functions:

- Transfer income and housing expenditure shares (s_t^L, s_h^L, s_h^H)
- Labor demand parameters (α, ρ, π)
- Transfer payment elasticity (Ψ)
- Mobility cost functions ($c^L(\cdot)$ and $c^H(\cdot)$)
- Housing supply function ($\Delta H^s(\cdot)$)

In order to reduce the number of parameters to estimate, I first impose values of s_t^L, s_h^L, s_h^H based on external information. I compute $s_t^L = 0.05$ by dividing

[36] These equations can be derived formally by including error terms which proportionally shift production, housing demand, housing supply, transfer payments, and indirect utility. For example, re-define the equilibrium condition for transfer payments as follows: $t_{it}^{L} = e_{it}^{t} \cdot \bar{T}^{L}\left(w_{it}^{L}\right)^{\Psi^{L}}$, where e_{it}^{t} is a random variable which represents unobservable shocks to transfer payment expenditures (and $E[e^t] = 1$). Totally differentiating this condition gives the following expression: $\Delta t_{it}^{L} = \Psi^{L}\left(\Delta w_{it}^{L}\right) + \Delta e_{it}^{t}$, which is the equation used in the GMM estimation.

aggregate expenditures on Food Stamps and Income Maintenance Programs by the sum of these expenditures and aggregate low-skill wage income. For the housing expenditure shares, I use $s_h^L = 0.34$ for non-college households and $s_h^H = 0.30$ for college-educated households based on the data presented in Section 2.[37]

For the labor demand curve, I compute $\pi = 0.37$ based on average wages for high-skill and low-skill workers and average share of high-skill workers in the adult population.[38] I choose $\rho = 0.29$ based on Katz and Murphy (1992).[39] This leaves the returns to scale parameter (α) to be estimated. Although this parameter will be estimated from functional form assumptions, it is still useful to include the two moments of the labor demand curve to check the overall fit of the model.[40] I also report results below which drop the labor demand moments.[41]

Finally, I choose the following functional forms for the mobility cost functions and housing supply elasticity:

$$c^j(x) = \frac{\sigma^j(\exp(\beta^j x) - 1)}{\beta^j} \quad j \in \{L, H\}$$
$$\Delta H^s(x) = \frac{\sigma^h(\exp(\beta^h x) - 1)}{\beta^h}$$

These functions are the exponential transformations suggested by Manly (1976),

[37]Average household income is \$82,439 for high-skill households in the baseline sample and is \$48,456 for low-skill households. Assuming $s_h = 0.30$ for high-skill households and income elasticity of 0.8, then $s_h = 0.34$ for low-skill households.

[38]I compute the wage premium (ζ) as 1.75, which is the average wages of college-educated workers divided by the average wages of non-college workers. I next compute the average share (over this time period) of college-educated workers in the labor force (μ) as 0.25. Using the formula for π in Section 2, this gives $\pi = 0.45$. The aggregate housing demand share parameter is given by $\nu = \mu\zeta s_h^H/(\mu\zeta s_h^H + (1-\mu)s_h^L)$.

[39]Katz and Murphy (1992) estimate the elasticity of substitution between high-skill and low-skill labor ($\sigma_{H,L}$) to be 1.4. This gives $\rho = 1 - 1/\sigma_{H,L} = 0.29$.

[40]Since the instrumental variable shifts the labor demand curve, parameters of the labor demand curve are identified from functional form assumptions.

[41]Because the labor demand instrument is measured with error, when using it in the GMM estimation, I rescale it by regressing adjusted wages on the instrument and scale the instrument so that this regression with the rescaled instrument would give a coefficient of 1.0. A more rigorous alternative is to modify the labor demand moments to include an additional parameter (κ) as follows:

$$\Delta e_{it}^{wH} = \Delta w_{it}^H - (\kappa\Delta\theta_{it} + ((\rho-1) + (\alpha-\rho)(\pi))\,\Delta H_{it} + (\alpha-\rho)(1-\pi)\Delta L_{it})$$
$$\Delta e_{it}^{wL} = \Delta w_{it}^L - (\kappa\Delta\theta_{it} + ((\rho-1) + (\alpha-\rho)(1-\pi))\Delta L_{it} + (\alpha-\rho)(\pi)\Delta H_{it})$$

This procedure yields very similar results.

which represent Box-Cox transformations of exponentiated variables and are defined so that if $\beta^j = 0$, then the functions simplify to $\sigma^j x$. These functions are flexible enough to accommodate interesting curvature with only two parameters, and they are everywhere monotonic and have continuous first derivatives, which greatly simplifies the computation. Ultimately, there are eight remaining parameters to estimate: $\{\sigma^L$, β^L, σ^H, β^H, σ^h, β^h,Ψ, $\alpha\}$.

The resulting GMM estimator solves a nonlinear, simultaneous equations problem, so in order to estimate the nonlinear parameters I need to take nonlinear functions of the instrumental variable ($\Delta\theta$) to achieve identification. I use $\Delta\theta$, $(\Delta\theta)^2$, $(\Delta\theta)^3$, $(\Delta\theta)^4$, and $(\Delta\theta)^5$ as instrumental variables. This results in 30 moment conditions (the five polynomial functions of the instrument $\times$ the six error terms). The full model is estimated using a standard two-step optimal GMM procedure (see Appendix for details of this procedure).

The GMM estimates are presented in Table 11. The first row presents the preferred specification using the external estimates discussed above. Columns (1) and (2) report estimates of the housing supply curve. The estimates suggest that the housing supply curve is concave ($\beta^h = 3.773$, s.e. 1.546). One way to interpret the housing supply coefficients is to compute the increase in housing supply when housing prices exogenously rise by 10% (28.3%) and compare it to the decrease in housing supply when housing prices decline by 10% (−19.4%). In other words, the magnitude of housing supply response is about 46% larger for a one standard deviation increase in housing prices than the response to a one standard deviation decrease.

The estimates of the mobility cost function parameters (columns (3) through (6)) give no evidence of an asymmetric mobility cost function for either high-skill or low-skill workers; the estimates suggest that the mobility cost functions are approximately linear. The point estimates for σ^L and σ^H are precisely estimated and statistically significantly different from zero, suggesting the existence of non-negligible mobility costs. To get a sense of the magnitudes, the point estimates imply that the 10th percentile of mobility costs in a city (i.e., the marginal migrant after 10% of the population has out-migrated following a negative shock) is roughly 24.9% of annual income for high-skill workers and 18.1% of annual income for low-skill workers.[42] In other words, despite the fact that low-skill workers are disproportionately likely to remain in declining cities following negative shocks, the point estimates imply that high-skill workers have larger mobility costs, in both relative as well as absolute terms (though the difference in relative terms is not large or precisely estimated). Column (8) reports the estimated transfer payment elasticity, which is quantitatively large and precisely estimated; the coefficient implies that a 1% decline in low-skill wages increases transfer payment expenditures by 4.5%. Column (9) reports estimates of the returns to scale parameter ($\alpha = 1.040$, s.e. 0.020), which suggests that returns to scale are approximately constant; this is also consistent with the reduced form results, which found no evidence of an asymmetric response of wages.[43] Lastly, the results in column (10) show that the overidentification test does not reject the null hypothesis that the deviations of the empirical moments from the model are due to chance ($p = 0.707$).

The remainder of Table 11 reports estimates of the full model under alternative economic assumptions. The second row reports estimates assuming that both housing expenditure share and public assistance expenditures do not differ by skill and are negligible (i.e., $s_h^H = s_h^L = 0.001$ and $s_t^H = s_t^L = 0.001$). These estimates verify that ignoring the welfare effects of housing price adjustments and changes in expenditures on public assistance programs results in much larger estimates of mobility costs for both high-skill and low-skill workers. In this scenario, the mobility cost estimates for low-skill workers are significantly larger in magnitude ($\sigma^L = -0.299$ versus $\sigma^H = -0.188$). Also, the difference between these coefficients is highly significant ($p < 0.001$). To compare to the baseline estimates, the mobility costs are roughly four times larger for low-skill workers and two times larger for high-skill workers when ignoring housing costs and transfer payments.[44]

[42] I assume the marginal migrant has 25 years of working life remaining and thus must trade off remaining to face permanently lower wage and employment opportunities against paying the one-time mobility cost to out-migrate and avoid the adverse wage and employment consequences.

[43] Wages did not respond asymmetrically but population and employment did, which suggests constant returns to scale. If there were decreasing returns to scale, then the asymmetric response of employment to the local labor demand shock would imply an asymmetric wage response, as well.

[44] The estimated mobility cost functions are also statistically significantly convex, implying that

The third and fourth rows report model estimates when only housing and only transfers are "shut down", respectively. The estimated mobility cost functions from these rows and the first two rows are graphed in Figure 6. Both the figure and the model estimates (see column (7) of row 4) suggest that transfer payments are responsible for a majority of the relative difference in mobility by skill. However, the magnitudes of mobility cost estimates are much larger for both types of workers when housing expenditures are ignored. In other words, the asymmetric population response for both high-skill and low-skill workers to exogenous labor demand shocks is primarily due to the asymmetric housing supply curve.

Rows 5 and 6 in Table 11 report estimates which impose alternative values of $\sigma_{H,L}$. First, I impose $\sigma_{H,L} = 50$, which corresponds to the two types of labor being close to perfect substitutes. The next row imposes $\sigma_{H,L} = 0.05$, which corresponds to the two types being close to perfect complements. In both cases, the estimates of the housing supply curve and mobility costs are not greatly affected; however, comparing these two rows to the baseline (row 1), it is worth noting that the fit of the model is best when using $\sigma_{H,L} = 1.4$ as opposed to the other extreme values.

The next row of Table 11 (row 7) uses an alternative measure of wages. As discussed above, the preferred measure of wages ("adjusted wages") assumes that most of the observed change in labor force participation is involuntary. This measure was chosen to provide an upper bound of estimated mobility costs. As an alternative, row 7 reports results using the "residualized wage" measure (see Section 4 for definition). Since residualized wages do not account for changes in labor force participation, the estimated mobility cost parameters are much lower. In fact, for low-skill workers I cannot reject the null hypothesis that mobility costs are zero. Overall, I conclude that these results suggest that mobility costs for both high-skill and low-skill workers are at most modest. Even under the extreme assumption that reservation wages are negligible, the estimated mobility costs are still much lower than would be implied by focusing solely on wages.[45]

the mobility cost of the marginal out-migrant rises faster than the marginal in-migrant, although the magnitude of the convexity is not large.

[45]The final row reports estimates which drop the labor demand curve moments. The reason why

One use of the GMM estimates is to construct out-of-sample counterfactual simulations of alternative policies towards social transfers. Figure 7 reports results from one such simulation. In this simulation, the system of means-tested transfers (summarized by the parameter Ψ) has been replaced by a system of mobility subsidies which reduces the mobility costs of all workers by 50%.[46] Each panel in the figure shows the response of a different endogenous variable. The figure shows that the mobility subsidies increase magnitude of low-skill out-migration following adverse shocks relative to the system of means-tested transfer payments. Therefore, the high-skill population share is much less responsive to shifts in local labor demand with mobility subsidies. One motivation for such a policy would be if there are strong negative externalities from increasing concentrations of low-skill workers in a particular area; in this case, mobility subsidies appear to provide consumption smoothing to low-skill workers without reducing their incentive to out-migrate.

1.6 Conclusion

Low-skill workers are comparatively immobile. When labor demand slumps in a city, college-educated workers tend to relocate whereas non-college workers are disproportionately likely to remain to face declining wages and employment. These facts may indicate that mobility is disproportionately costly for low-skill workers. This paper proposes and tests an alternative explanation, which is that the incidence of adverse labor demand shocks is borne in large part by (falling) real estate rental prices and (rising) social transfers. The spatial equilibrium model developed in this paper illustrates how wages, employment, population, housing prices, and transfer payments re-equilibrate after a local labor demand shock. Appropriately parameterized, this

alternative assumptions on the elasticity of substitution did not substantially affect the estimated mobility cost functions is that the labor demand moments contribute to identification only indirectly through the optimal GMM weighting matrix estimated in the first step of the two-step procedure. Therefore, it is not surprising that dropping the labor demand moments entirely does not significantly affect the estimates of the mobility cost functions (Table 11, row 8).

[46] Although this is an obviously stylized form of mobility subisidies, it is not an unrealistic approximation if the policy took the form of a tax credit that was indexed to income. Recall that mobility costs in the model are defined as a fraction of annual income.

model identifies both the magnitude of unobserved mobility costs by skill and the shape of the local housing supply curve.

Using U.S. Census data, nonlinear reduced form estimates of the effect of plausibly exogenous labor demand shocks document that positive labor demand shocks increase population more than negative shocks reduce population, that this asymmetry is larger for low-skill workers, and that such an asymmetry is absent for wages, housing values, and rental prices.

These facts are consistent with the presence of limited mobility costs for high-skill and low-skill workers and a concave housing supply curve (most likely due to a durable housing stock). Estimates of a full spatial equilibrium model using a nonlinear, simultaneous equations GMM estimator are consistent with the reduced form evidence and suggest that the primary explanation for the comparative immobility of low-skilled workers is not higher mobility costs per se, but rather a lower incidence of adverse local demand shocks.

The finding that mobility costs are limited for both high-skill and low-skill workers is a necessary condition to be able to properly interpret changes in housing values due to changes in observed local amenities as a valid marginal willingness to pay for the amenity (see, for example, Chay and Greenstone (2003)). The results in this paper suggest that the assumption of perfect mobility may be a valid approximation in some of these hedonic studies, especially when evaluating changes in local amenities over decadal time horizons.[47]

One important area of future work is how the incidence of local labor market shocks is shared between homeowners and renters. On the one hand, homeowners' "user cost" of housing has declined following a negative labor demand shock; on the other hand, however, declines in housing values have a negative wealth effect which may affect how responsive the household is to local labor demand shocks. A full

[47]It is worth stressing that even over decadal time horizons the assumption of perfect mobility is only an approximation. The preferred GMM estimates in this paper imply non-negligible magnitudes of mobility costs for both high-skill and low-skill workers following large negative shocks, suggesting that it may be appropriate to estimate hedonic models which incorporate barriers to migration when the underlying changes in local amenities are large. See Bayer et al (2008) for work in this direction.

assessment of the incidence of local labor market shocks thus awaits further study. Another important area of future work is looking at individual transfer programs. For example, the federal disability insurance program rules suggest that the take-up decision is generally a once-and-for-all choice, so that disability insurance receipt is an absorbing state (Autor and Duggan, 2003). The econometric setup in this paper could be used to test whether positive shifts in local labor demand increase DI takeup by less than negative shifts increase DI takeup.

Lastly, the finding that mobility costs are limited suggests that transfer payments may be significantly crowding out the individual migration decision for low-skill workers, which is consistent with the results in the recent "welfare magnetism" literature (see, for example, Gelbach (2004)). This implies that the social efficiency of public insurance programs may depend on the geographic breadth of an adverse labor demand shock, since when a shock is geographically broad (as during a recession), the gains to relocation are small and there is less scope for transfer payments to crowd out migration.

1.7 References

Abraham, Katherine and John C. Haltiwanger. 1995. "Real Wages and the Business Cycle." *Journal of Economic Literature,* 33(3): 1215-1264.

Albouy, David. 2009a. "The Unequal Geographic Burden of Federal Taxation." *Journal of Political Economy*, 117(4): 635-667.

Albouy, David. 2009b. "What are Cities Worth? Land Rents, Local Productivity, and the Capitalization of Amenity Values." NBER Working Paper #14981.

Artuc, Erhan, Shubham Chaudhuri, and John McLaren. 2007. "Trade Shocks and Labor Adjustment: A Structural Empirical Approach." *American Economic Review.*

Autor, David and David Dorn. 2009. "Inequality and Specialization: The Growth of Low-Skill Service Jobs in the United States," mimeo.

Autor, David and Mark Duggan. 2006. "The Rise in Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics*, 118(1): 157-205.

Bayer, Patrick, Nathanial Keohane, and Christopher Timmins. 2008. "Migration and Hedonic Valuation: The Case of Air Quality." *Journal of Environmental Economics and Management*, forthcoming.

Bartik, Timothy J. 1991. "Who Benefits From State and Local Economic Development Policies." W.E. Upjohn Institute for Employment Research: Kalamazoo, Mich.

Blanchard, Olivier Jean and Larry Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity*, 1-75.

Bound, John and Harry J. Holzer. 2000. "Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s." *Journal of Labor Economics*, 18(1): 20-54.

Chaudhuri, Shubham and John McLaren. 2007. "Some Simple Analytics of Trade and Labor Mobility." NBER Working Paper #13464.

Chay, Kenneth Y. and Michael Greenstone. 2005.. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy*, 113(2): 376-424.

Chernozhukov, Victor, Ivan Fernadez-Val and Alfred Galichon. 2009. *Biometrika*, forthcoming.

Ellison, Glenn and Sara Fisher Ellison. 2000. "A simple framework for nonparametric specification testing." *Journal of Econometrics*, 96(1): 1-23.

Gallin, Josh. 2004. "Net Migration and Labor Market Dynamics." *Journal of Labor Economics*, 22: 1-21.

Gelbach, Jonah. 2004. "Migration, the Lifecycle, and State Benefits: How Low is the Bottom?" *Journal of Political Economy.* 112(5), 1091-1130.

Glaeser, Edward L. and Joseph Gyourko. 2005. "Urban Decline and Durable Housing." *Journal of Political Economy*, 113(2): 345-375.

Greene, William H. 2003. *Econometric Analysis.* Fifth Edition, Prentice Hall: New Jersey.

Gould, Eric, David B. Mustard, and Bruce Weinberg. 2002. "Crime Rates and Local Labor Market Opportunities in the United States: 1979-1997." *Review of Economics and Statistics.* 84(1): 45-61.

Hall, Peter. 1992. "Effect of Bias Estimation on Coverage Accuracy of Bootstrap Confidence Intervals for a Probability Density." *Annals of Statistics*, 20(2): 675-694.

Hansen, Lars. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 50: 1029-1054.

Horowitz, Joel. 2001. "The Bootstrap." *Handbook of Econometrics*, Volume 5.

Hoynes, Hillary. 1999. "The Employment, Earnings, and Income of Less-Skilled Workers over the Business Cycle." Institute for Research on Poverty Discussion Paper, no. 1199-99.

Jayachandran, Seema. 2006. "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries." *Journal of Political Economy*, 114(3): 2006.

Katz, Larry and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics*, 102(1): 35-78.

Kennan, John and James R. Walker. 2009. "The Effect of Expected Income on Individual Migration Decisions," mimeo.

Kennan, John and James R. Walker. 2008. "Wages, Welfare Benefits and Migration," mimeo.

Luttmer, Erzo. 2005. "Neighbors as Negatives: Relative Earnings and Well-Being." *Quarterly Journal of Economics*, 120(3): 963-1002.

Malamud, Ofer and Abigail Wozniak. 2008. "The Impact of College Graduation on Geographic Mobility: Identifying Education Using Multiple Components of Vietnam Draft Risk." Working Paper.

Manly, B. F. J. 1976. Exponential data transformations. *Journal of the Royal Statistical Society, Series D (The Statistician).* 25(1), 37-42.

Meyer, Bruce, Wallace K. C. Mok, and James X. Sullivan. 2009. "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences." Harris School Working Paper #09.03.

Morretti, Enrico. 2004. "Estimating the Social Return to Higher Education: Evidence From Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics.* 121(1-2).

Morretti, Enrico. 2009. "Real Wage Inequality", mimeo.

Polinsky, A. Mitchell and David T. Ellwood. 1979. "An Empirical Reconciliation of Micro and Macro Grouped Estimates of the Demand for Housing." *The Review of Economics and Statistics*, 61(2): 199-205.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy*, 90(6): 1257-1278.

Ruggles, Steven, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. 2004. *Integrated Public Use Microdata Series: Version 3.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor].

Saiz, Albert. 2009. "On Local Housing Supply Elasticity." Working Paper.

Saks, Raven. 2005. "Job Creation and Housing Construction: Constraints on Metropolitan Area Employment Growth." Finance and Economics Discussion Series 2005-49. Washington: Board of Governors of the Federal Reserve System.

Shapiro, Jesse. 2006. "Quality of life, productivity, and the growth effects of human capital." *Review of Economics and Statistics.*

Topel, Robert H. 1986. "Local Labor Markets." *Journal of Political Economy*, 94(3): S111-S143.

Wozniak, Abigail. 2006. "Why Are College Graduates More Responsive to Distant Labor Market Opportunities?" Working Paper.

1.8 Appendix

1.8.1 Data Appendix

U.S. Census Data

The sample of adults used in the analysis includes all individuals between the age of 18 and 64, were not in group quarters such as prisons and psychiatric institutions, and who lived in a metropolitan area available in the Census IPUMS. All available MSAs are used in analysis except for Biloxi-Gulfport, MS, Flint, MI, and Reno, NV. These MSAs are dropped because of obvious mismeasurement of the labor demand shock. Specifically, in at least one of the decades in the sample, these MSAs experienced a greater than one standard deviation labor demand shock according to the predicted labor demand instrument but experienced a greater than one standard deviation change in population and rental prices of the opposite magnitude. All results including these cities are similar to the main results in Tables 2 through 4.

Individuals are dropped if they report business income, farm income or work in farming or agriculture. Individual labor supply is measured by multiplying weeks worked times usual weekly hours worked. To be included in the sample of workers used to construct the predicted employment measure, the worker must be in the labor force and have positive and non-missing hours worked and annual income.

Individual hourly wages are computed by dividing yearly wage and salary income by the product of weeks worked and usual weekly hours worked. Topcoded yearly wage income values are multiplied by 1.5 and (following Autor and Dorn (2009)) hourly wages are set not to exceed this value divided by (50 weeks $\times$ 35 hours). Local area wage statistics are computed based on the sample of workers who work at least 35 weeks and at least 30 hours per week. Wages are deflated using the CPI-U series.

In order to construct an estimate of the local area wage premium, log wages of the sample described above are regressed on MSA fixed effects, a quadratic in potential experience (age $-$ years of education $-$ 6), 14 industry dummy variables, 6 occupation category dummy variables, and dummy variables for gender, veteran status, marital status, and race. This regression is run each decade and in each decade is run separately for workers with and without a college degree. In each case, the magnitude of the MSA fixed effects corresponds to the local area wage premium. All regressions and calculations of local area averages are computed using the Census individual sampling weights.

The rental price and housing value local area premiums are computed similarly to the wage premiums; namely, I regress the log of these variables on a quadratic in the number of bedrooms and the number of rooms and an interaction between number of bedrooms and number of rooms. These regressions and calculations of (unconditional) average rental prices and housing values use the Census household weights since the housing value and rental price data are reported at the household level. Topcoded rental prices and housing values are multiplied by 1.5.

Regional Economic Information System (REIS)

The REIS data come from the Bureau of Economic Analysis.[48] I aggregate the county-level data into MSAs using the 1990 MSA definitions. When a county spans multiple MSAs I use 1990 population weights to assign fractions of the county totals across the various MSAs.

County Business Patterns (CBP)

The County Business Patterns data are available from the U.S. Census Bureau and the ICPSR data repository.[49] I used the 1979, 1989, and 1997 CBP data to match the 1980, 1990, and 2000 Census data described above. The 1997 CBP data were chosen because the 1998 and 1999 CBP data use the NAICS industry codes, while the CBP data before 1997 used SIC codes. I use 3-digit SIC industry codes to construct the alternative measure of predicted employment. Roughly 35 percent of the county-by-industry employment cells are suppressed. In these cases, I observe the number of establishments in each establishment size bin and a flag indicating the range of actual

[48]See this website for more information: http://www.bea.gov/regional/reis/default.cfm#step2.

[49]I downloaded the 1989 and 1997 CBP data from the following URL: http://www.census.gov/econ/cbp/historical.htm. The 1979 CBP data were downloaded from ICPSR at the following URL: http://www.icpsr.umich.edu/icpsrweb/ICPSR/series/00022.

employment. To compute predicted employment for these cells, I run a regression each year using the non-suppressed data and use this regression to compute predicted employment for suppressed cells from the fitted values. I then compare total county employment from the raw CBP data to the total county employment computed using the non-suppressed cells and the predicted employment values. If these two values are not within 1%, then I scale all of the predicted employment values by a scalar so as to make the two totals equal, and I then check again that the predicted values lie within the ranges indicated by the employment flag and I continue to repeat this procedure until the two totals are within 1%.

1.8.2 Comparative Statics

This subsection derives comparative statics for the model described in Section 2 in the special case when there are constant returns to scale ($a = 1$). The comparative statics are derived for the following three scenarios:

- Case 1: No mobility costs; constant housing supply elasticity
- Case 2: No mobility costs; concave housing supply curve
- Case 3: Large mobility costs; constant housing supply elasticity

Case 1: No mobility costs; constant housing supply elasticity

This case corresponds to the following restrictions on the housing supply curve and the mobility cost functions: $c^L(\Delta L_{it}) = 0$, $c^H(\Delta H_{it}) = 0$, and $\Delta H^S(\Delta p^h_{it}) = \sigma \cdot \Delta p^h_{it}$. With no mobility costs and a constant housing supply elasticity, the model readily admits a closed-form solution. Additionally, in this case all endogenous variables respond *symmetrically* – meaning that equal-sized positive and negative exogenous labor demand shocks cause positive and negative shifts of equal magnitude in all of the endogenous variables (Δw^H_{it}, Δw^L_{it}, ΔH_{it}, ΔL_{it}, Δp^h_{it}, Δt_{it}). Mathematically, this means that changes in the endogenous variables are linear functions of the exogenous labor demand shocks; i.e., $\Delta x_{it} = \bar{K}^x \Delta \theta_{it}$ where x is one of the endogenous variables in the model. To derive these results, first note that with no mobility costs, housing prices will respond symmetrically to both high-skill and low-skill wages:

$$\Delta p_{it}^h = \frac{1}{s_h^H} \Delta w_{it}^H$$
$$\Delta p_{it}^h = \left(\left((1 - s_t^L) + s_t^L \Psi^L \right) / s_h^L \right) \Delta w_{it}^L \equiv \left(\Gamma^L / s_h^L \right) \Delta w_{it}^L$$

Next, note that with constant returns to scale, wages for high-skill and low-skill workers can be written as follows:

$$\pi \cdot \Delta w_{it}^H + (1 - \pi) \cdot \Delta w_{it}^L = \Delta \theta_{it}$$

Combining the three previous expressions gives the following:

$$\Delta w_{it}^H = \left(s_h^H \Gamma^L / \left(\pi s_h^H \Gamma^L + (1 - \pi) s_h^L \right) \right) \Delta \theta_{it} \equiv \bar{K}^{wH} \cdot \Delta \theta_{it}$$
$$\Delta w_{it}^L = \left(s_h^L / \left(\pi s_h^H \Gamma^L + (1 - \pi) s_h^L \right) \right) \Delta \theta_{it} \equiv \bar{K}^{wL} \cdot \Delta \theta_{it}$$
$$\Delta p_{it}^h = \left(\Gamma^L / \left(\pi s_h^H \Gamma^L + (1 - \pi) s_h^L \right) \right) \Delta \theta_{it} \equiv \bar{K}^{p} \cdot \Delta \theta_{it}$$

In other words, with no mobility costs for workers and firms, wages and housing prices respond symmetrically. Transfer payments will also respond symmetrically since transfer payments are a log-linear function of low-skill wages: i.e., $\Delta t_{it}^L = \Psi^L \cdot \bar{K}^{wL} \cdot \Delta \theta_{it} \equiv \bar{K}^t \cdot \Delta \theta_{it}$.

Finally, with a constant housing supply elasticity, the migration response is also symmetric, since ΔH_{it} and ΔL_{it} can be written as linear functions of Δw_{it}^H, Δw_{it}^L, and Δp_{it}^h. Simple algebra gives the following two expressions:

$$\Delta H_{it} = \bar{K}^H \Delta \theta_{it}$$
$$\Delta L_{it} = \bar{K}^L \Delta \theta_{it}$$

where $\bar{K}^H$ and $\bar{K}^L$ are constants that can be written in terms of the primitive parameters of the model (α, ρ, π, s_h^H, s_h^L, s_t^L).[50] In summary, the log-linearity of the housing supply curve and the absence of mobility costs implies that all endogenous variables respond symmetrically to the exogenous labor demand shock.

Case 2: No mobility costs; concave housing supply curve

Formally, this case can be written as follows: $c^L(\Delta L_{it}) = 0$, $c^H(\Delta H_{it}) = 0$, and $\Delta H^S(\Delta p_{it}^h)$ is increasing in Δp_{it}^h. As in the previous case (and following the same derivation), wages, housing prices, and transfer payments all respond symmetrically, with the same constant terms as above:

$$\Delta w_{it}^H = \bar{K}^{wH}\Delta\theta_{it};\ \Delta w_{it}^L = \bar{K}^{wL}\Delta\theta_{it};\ \Delta p_{it}^h = \bar{K}^p\Delta\theta_{it};\ \Delta t_{it}^L = \bar{K}^t\Delta\theta_{it}$$

In case 2, however, population no longer responds symmetrically to the exogenous shock. To see this, go back to the housing market equilibrium condition and substitute the expressions above:

$$\bar{K}^p\Delta\theta_{it} + \Delta H^S(\bar{K}^p\Delta\theta_{it}) = (\bar{K}^{wH} + (1 - s_t^L)\bar{K}^{wL} + s_t^L\bar{K}^t)\Delta\theta_{it} + \Delta H_{it} + \Delta L_{it}$$

Since the elasticity of substitution between high-skill and low-skill labor is constant, we know that $\Delta w_{it}^H - \Delta w_{it}^L = (\rho - 1)(\Delta H_{it} - \Delta L_{it})$. Combining these two expressions gives the following expressions for ΔH_{it} and ΔL_{it}:

$$\Delta H_{it} = \frac{1}{2}(\Lambda^H\Delta\theta_{it} + \Delta H^S(\bar{K}_1^p\Delta\theta_{it}))$$
$$\Delta L_{it} = \frac{1}{2}\left(\Lambda^L\Delta\theta_{it} + \Delta H^S(\bar{K}_1^p\Delta\theta_{it})\right)$$

[50] The constants $\bar{K}^H$ and $\bar{K}^L$ are defined as follows:

$$\bar{K}^H = \frac{(\bar{K}^{wH} - \bar{K}^{wL}) - (\rho - 1)(\bar{K}^{wH} + \Gamma\bar{K}^{wL} - \bar{K}^p(1+\sigma))}{2(\rho - 1)s_h^H}$$
$$\bar{K}^L = \bar{K}^H - (\bar{K}^{wH} - \bar{K}^{wL})/(\rho - 1)$$

where Λ^H and Λ^L are constant terms.[51] Since $\Delta H^S(x)$ is increasing in x, these expressions imply that ΔH_{it} and ΔL_{it} are increasing in $\Delta\theta_{it}$. In other words, because the housing supply curve is concave, the responsiveness of high-skill and low-skill population to exogenous shocks is convex – positive local labor demand shocks increase population more than negative shocks reduce population. It is also possible to show that if $s_h^H < s_h^L < 1$ and $\Psi^L < 0$, and $s_t^L > 0$, then decreases in local labor demand will reduce the fraction high-skill workers in the population.[52]

Case 3: Large mobility costs; constant housing supply elasticity

Formally, this case can be defined as follows: $c^L(\Delta L_{it})$ and $c^H(\Delta H_{it})$ are declining and convex functions and $\Delta H^S(\Delta p_{it}^h) = \sigma \cdot \Delta p_{it}^h$. In this case, the convexity of the mobility cost functions imply that the mobility cost of the marginal migrant is greater in magnitude for decreases in population than for equal-sized increases in population. As mentioned in the introduction, one way this could arise is if the city is small relative to the rest of the world, so that the mobility cost of the marginal in-migrant is negligible. In this case, $c^H(\Delta H_{it})$ would be defined such that $c^H(\Delta H) = 0$ for all $\Delta H_{it} \geq 0$, but $c^H(\Delta H_{it})$ is decreasing in ΔH_{it} for all $\Delta H_{it} < 0$.

In the case where high-skill and low-skill labor differ only in productivity (i.e., $s_t^H = s_t^L = s_t$, $s_h^H = s_h^L = s_h$, $\Psi^H = \Psi^L = \Psi$, and $c^L(x) = c^H(x)\ \forall x$), it can be shown that wages still respond symmetrically as in the previous two cases. By simplifying the problem to make high-skill and low-skill workers identical except for their efficiency units of labor, it is straightforward to show that $\Delta w_{it}^H = \Delta w_{it}^L = \Delta\theta_{it}$ and that $\Delta H_{it} = \Delta L_{it}$. These simplifications result in the following expressions for

[51] The constants Λ^H and Λ^L are defined as follows:

$$\Lambda^H = \bar{K}^p + \bar{K}^{wH} + \bar{K}^{wL} + (\bar{K}^{wH} - \bar{K}^{wL})/(\rho - 1)$$
$$\Lambda^L = \Lambda^H - 2(\bar{K}^{wH} - \bar{K}^{wL})/(\rho - 1)$$

[52] To see this, note that $\Delta H - \Delta L = \left(K^{wH} - K^{wL}\right)/(\rho - 1) \cdot \Delta\theta$. If $s_h^L > s_h^H$, $\Psi^L < 0$, and $s_t^L > 0$, then $\left(K^{wH} - K^{wL}\right) < 0$. Since $\rho - 1 < 0$ (since $\sigma_{H,L} > 0$), this implies $\left(K^{wH} - K^{wL}\right)/(\rho - 1) > 0$. Thus declines in $\Delta\theta$ will reduce $\Delta H - \Delta L$.

housing market and labor supply conditions, respectively:

$$\Delta p_{it}^{h} \cdot (1+\sigma) - 2s_h(\Gamma\Delta\theta_{it} + \Delta H_{it}) = 0 \tag{1.8}$$
$$c(\Delta H_{it}) + \Gamma\Delta\theta_{it} - s_h\Delta p_{it}^{h} = 0$$

where $\Gamma = (1-s_t) + s_t\Psi$. Combining these expressions gives the following:

$$2s_h\Delta H_{it} - (1+\sigma)c(\Delta H_{it}) = (1+\sigma-2s_h)\Gamma\Delta\theta_{it}$$

Since $c(\Delta H_{it})$ is declining and convex, this implies that ΔH_{it} is convex in $\Delta\theta_{it}$.[53] Since ΔH_{it} is convex in $\Delta\theta_{it}$, then by equation (1.8), Δp_{it}^{h} is convex in $\Delta\theta_{it}$. In other words, unlike the other cases, in this case housing prices respond asymmetrically, where positive shocks increase housing prices more than negative shocks reduce housing prices. The intuition is that the convexity of mobility cost function makes out-migration disproportionately costly (as compared to in-migration). Because of these mobility costs, following negative shocks workers are willing to pay more for housing than they would in the absence of mobility costs, which bids up the price of housing following price declines.

1.8.3 Model Simulation Details

The data used to create Figure 3 are simulated from the model described in Section 2. The same parameters used in the GMM estimation are used in the simulation; i.e., $s_t^L = 0.05$, $s_h^L = 0.34$, $s_h^H = 0.30$, $\rho = 0.29$, $\pi = 0.45$. The returns-to-scale parameter $\alpha = 1$ is used, and the transfer payment elasticity used is $\Psi^L = -5.0$. The mobility cost functions and housing supply function are parameterized as they are in the GMM estimator: i.e., $c^L(x) = \sigma^L(\exp(\beta^L x) - 1)/\beta^L$, $c^H(x) = \sigma^H(\exp(\beta^H x) - 1)/\beta^H$, $\Delta H^s(x) = \sigma^h(\exp(\beta^h x) - 1)/\beta^h$. The values of these parameters depend on the

[53]Formally, a sufficient condition for this result to hold is that $(1-s_t) + s_t\Psi > 0$ and $s_h < ((1-s_t) + s_t\Psi)(1+\sigma)$. In words, transfer payments provide partial wage insurance, and housing expenditure share cannot be so large so that negative shocks would cause net in-migration of low-skill labor.

scenario as follows:

- Case 1: $\sigma^H = \sigma^L = 0$, $\sigma^h = 4.0$, $\beta^h = 0$
- Case 2: $\sigma^H = \sigma^L = 0$, $\sigma^h = 2.0$, $\beta^h = 4.0$
- Case 2: $\sigma^H = \sigma^L = -0.2$, $\beta^H = \beta^L = -100$, $\sigma^h = 4.0$, $\beta^h = 0$

1.8.4 A Simple Model of Durable Housing

This section outlines a model to provide simple microfoundations for a concave housing supply curve (i.e., housing supply elasticity that is larger for increases in housing demand than for decreases in housing demand). As in Section 2, the model here is a two period model, where a single city is shocked out of a large number of cities. The model includes a labor market and a housing market. Production of a homogeneous tradable good is constant returns to scale and uses only (homogeneous) labor as an input. All workers have identical Cobb-Douglas preferences for housing and the tradable good, so that expenditure share on housing (s_h) is constant.

Housing is supplied by absentee landlords who live in other cities. The housing supply is homogeneous in terms of workers' willingness-to-pay but there are heterogeneous costs to supplying housing (arising, perhaps, from topographic features of the land). This is modeled by assuming that the maximum housing supply is $\bar{H}^S$ (where $\bar{H}^S$ is assumed to be large enough so that we are not close to a corner solution) and that the cost of supplying an infinitesimal unit of housing is distributed according to the following density function: $f(c) = \frac{\sigma^h}{\bar{c}} \left(c/\bar{c}\right)^{\sigma^h - 1}$, where c is drawn from the closed interval $[0, \bar{c}]$. This results in an aggregate housing supply curve of $H^s(p^h) = \int_0^{p^h} \bar{H}^S f(c)dc = \bar{H}^S \cdot \left(p^h/\bar{c}\right)^{\sigma^h}$. Thus the initial housing market equilibrium is given by the following supply-demand equilibrium condition: $\bar{H}^S \cdot \left(p^h_{it}/\bar{c}\right)^{\sigma^h} = \bar{H}^D s_h w_{it} n_{it}/p^h_{it}$.

Using a similar simplifying assumption as in Glaeser and Gyourko (2005), I assume that housing is occasionally (and randomly) destroyed, and that the cost of rebuilding is the same as the initial cost of building. Mathematically, I assume that just before the labor demand shock, a random fraction δ of the initial housing supply collapses and needs to be re-built. For increases in housing demand, all housing that collapsed is immediately rebuilt in between periods, and housing supply further expands according to the elasticity of housing supply (σ^h). For decreases in housing demand, however, the "effective" housing supply elasticity is now only $\delta \cdot \sigma^h$ because some of the housing that was previously built does not collapse and cannot be destroyed. Unless $\delta = 1$ (i.e., housing is not durable at all and completely collapses between periods), these assumptions imply that the housing supply curve is nonlinear, asymmetric, and concave.

The equilibrium changes in wages, population, and housing prices following an exogenous labor demand shock ($\Delta\theta_{it}$) are as follows. The wage change in the city receiving the shock is $\Delta w_{it} = \Delta\theta_{it}$. Perfect mobility of workers implies that $\Delta w_{it} = s_h \Delta p_{it}^h$. This implies that $\Delta p_{it}^h = \Delta\theta_{it}/s_h$. In other words, both wages and housing prices respond *symmetrically.* For positive labor demand shocks, population increases by $\Delta n_{it} = (1+\sigma^h)\Delta p_{it}^h - \Delta w_{it} = (1+\sigma^h - s_h)/s_h \cdot \Delta\theta_{it}$. For negative labor demand shocks, population decreases by $\Delta n_{it} = (1+\delta\cdot\sigma^h - s_h)/s_h \cdot \Delta\theta_{it}$. Assuming $s_h > 0$, $\sigma^h > 0$ and $0 < \delta < 1$, then positive shocks increase population more than equal-sized negative shocks reduce population.

The key difference between this model and the model in Glaeser and Gyourko (2005) is that the marginal value and the average value of housing are equal in the simple model in this section, while in Glaeser and Gyourko (2005) housing units have heterogeneous, location-specific amenities, which causes average housing prices to respond asymmetrically due to compositional changes in the location-specific amenities in the housing stock. The empirical evidence in this paper suggests that housing prices respond symmetrically to exogenous labor demand shocks, which is more consistent with the model in this paper and suggests that location-specific amenities are not qualitatively important in determining equilibrium housing values and rental prices.

1.8.5 GMM Estimation

There are 30 empirical moments given by the following vector:

$$\mathbf{m} = (\mathbf{m}^1, \mathbf{m}^2, \mathbf{m}^3, \mathbf{m}^4, \mathbf{m}^5)'$$

where

$$\mathbf{m}^d = \begin{bmatrix} \Delta e^h(\Delta\theta)^d \\ \Delta e^{wH}(\Delta\theta)^d \\ \Delta e^{wL}(\Delta\theta)^d \\ \Delta e^H(\Delta\theta)^d \\ \Delta e^L(\Delta\theta)^d \\ \Delta e^t(\Delta\theta)^d \end{bmatrix}$$

The orthogonality conditions are summarized as $E[\mathbf{m}] = 0$. The parameters to estimate are given by the following vector:

$$\beta = (\sigma^h, \beta^h, \sigma^H, \beta^H, \sigma^L, \beta^L, \Psi, \alpha)'$$

The two-step GMM estimator is implemented by first estimating $\hat{\beta}^0$ as follows:

$$\hat{\beta}^0 = \arg\min_{\beta} \mathbf{m}'\mathbf{m}$$

This estimate is then used to form the following:

$$\hat{\Phi}^0 = \frac{1}{N}\sum_{i=1}^{N} \mathbf{m}_i(\hat{\beta}^0) \cdot \mathbf{m}_i'(\hat{\beta}^0)$$

Next, $\hat{\beta}$ is re-estimated as follows:

$$\hat{\beta}^{GMM} = \arg\min \mathbf{m}'(\hat{\Phi}^0)^{-1}\mathbf{m}$$

Inference is done by computing the following variance-covariance matrix:

$$\hat{V} = \frac{1}{N}\left(\hat{G}'(\hat{\Phi}^1)^{-1}\hat{G}\right)^{-1}$$

where $\hat{\Phi}^1$ is re-estimated using $\hat{\beta}^{GMM}$ instead of $\hat{\beta}^0$, and $\hat{G}$ is given by the following:

$$\hat{G} = \frac{1}{N}\sum_{i=1}^{N}\begin{bmatrix} \partial m_i^1/\partial\beta \\ \partial m_i^2/\partial\beta \\ \vdots \\ \partial m_i^{30}/\partial\beta \end{bmatrix}$$

Finally, the overidentification statistic is given by:

$$\mathbf{m}'(\hat{\beta}^{GMM}) \cdot (\hat{\Phi}^0)^{-1} \cdot \mathbf{m}(\hat{\beta}^{GMM}) \rightarrow \chi^2(\text{row}(\mathbf{m}) - \text{row}(\beta))$$

Table 1
Summary Statistics

Variable Name	N	Mean	Standard Dev.	Percentiles 5th	25th	50th	75th	95th
U.S. Census Data (IPUMS)								
Adult population (in millions)	645	0.425	0.856	0.060	0.093	0.177	0.392	1.477
Employment (in millions)	645	0.303	0.596	0.041	0.067	0.127	0.283	1.036
Employment-to-population ratio	645	0.711	0.051	0.625	0.680	0.714	0.748	0.786
Income per adult (in $000s)	645	14.979	3.167	10.516	12.871	14.664	16.674	20.079
Residualized wage ($)	645	11.545	1.207	9.801	10.718	11.399	12.304	13.712
Residualized wage, LFP adjusted ($)	645	8.225	1.131	6.593	7.496	8.142	8.911	10.095
College share of adult population	645	0.190	0.063	0.105	0.143	0.181	0.226	0.305
College share of employment	645	0.221	0.065	0.131	0.173	0.213	0.257	0.341
Average housing value (in $000s)	645	97.449	45.450	58.005	71.527	84.774	107.212	196.809
Average gross rent (in $000s)	645	5.229	1.014	4.055	4.579	5.017	5.581	7.196
REIS Data								
Food stamps + Income Maintenance (in $000s per non-college adult)	645	0.652	0.325	0.247	0.429	0.594	0.792	1.286

Notes: Baseline sample is a balanced panel of 215 Metropolitan Statistical Areas (MSAs), and all observations are MSA-year. IPUMS data are the 1980, 1990, and 2000. The REIS data are county-level and annual, but are aggregated to MSAs using the 1990 MSA definitions. All dollar values in this table are nominal, but all dollar-valued variables are converted to real dollars in the analysis. All specifications in subsequent tables are in first differences, so the three decades in this data set become two 10-year changes (thus, N=430 in the regressions that follow).

Table 2
Effects of Local Labor Demand Shocks

Dependent Variable:	(1) Population	(2) Employment	(3) Emp-to-Pop Ratio	(4) Income per 18-64 Adult	(5) Residualized Average Local Wage	(6) Residualized Wage, LFP Adjusted ("Adjusted Wage")
% Change in predicted employment (β)	1.800	2.054	0.089	0.959	0.353	0.520
	(0.445)	(0.466)	(0.038)	(0.137)	(0.086)	(0.109)
	[0.000]	[0.000]	[0.019]	[0.000]	[0.000]	[0.000]
(% Change in predicted employment)^2 (δ)	28.004	32.529	1.210	0.376	-0.753	1.454
	(7.898)	(9.096)	(0.797)	(2.858)	(1.640)	(2.422)
	[0.000]	[0.000]	[0.130]	[0.896]	[0.647]	[0.549]
Marginal effect at $-\sigma$	-0.115	-0.170	0.007	0.933	0.404	0.420
	(0.838)	(0.887)	(0.054)	(0.202)	(0.154)	(0.171)
	[0.891]	[0.848]	[0.905]	[0.000]	[0.009]	[0.015]
Marginal effect at $+\sigma$	3.716	4.279	0.172	0.984	0.301	0.619
	(0.528)	(0.648)	(0.076)	(0.271)	(0.128)	(0.222)
	[0.000]	[0.000]	[0.026]	[0.000]	[0.019]	[0.006]
p-value of nonparametric specification test	0.000	0.000	0.270	0.523	0.624	0.493
R^2	0.318	0.355	0.603	0.671	0.470	0.342
N	430	430	430	430	430	430

Notes: All columns report OLS results from estimating equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs. Dependent variable is always the percentage change across periods, except for column (3) where it is the percentage point change. The Residualized Wage in column (5) controls for observed compositional changes in the labor force between periods. The Adjusted Wage in column (6) uses the Residualized Wage and additionally accounts for changes in labor force participation. See text and Data Appendix for more details. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table 3
Effects of Labor Demand Shocks by Skill

Dependent Variable:	(1) Adult Population, College	(2) Adult Population, Non-College	(3) College Share of Adult Population	(4) Total Employed, College	(5) Total Employed, Non-College	(6) College Share of Employed	(7) Residualized Wage, College	(8) Residualized Wage, Non-College	(9) Adjusted Wage, College	(10) Adjusted Wage, Non-College
% Change in predicted employment (β)	1.923	1.607	0.051	2.194	1.858	0.039	0.297	0.345	0.467	0.513
	(0.544)	(0.436)	(0.024)	(0.553)	(0.458)	(0.027)	(0.080)	(0.085)	(0.099)	(0.107)
	[0.001]	[0.000]	[0.036]	[0.000]	[0.000]	[0.151]	[0.000]	[0.000]	[0.000]	[0.000]
(% Change in predicted employment)^2 (δ)	35.196	28.051	-0.816	36.970	32.886	-1.021	-1.319	-0.827	-0.498	1.520
	(10.359)	(7.685)	(0.349)	(10.988)	(8.829)	(0.376)	(1.397)	(1.625)	(2.008)	(2.367)
	[0.001]	[0.000]	[0.020]	[0.001]	[0.000]	[0.007]	[0.346]	[0.611]	[0.804]	[0.521]
Marginal effect at $-\sigma$	-0.484	-0.312	0.107	-0.335	-0.392	0.109	0.387	0.402	0.501	0.409
	(0.977)	(0.834)	(0.033)	(0.986)	(0.881)	(0.037)	(0.138)	(0.151)	(0.166)	(0.166)
	[0.621]	[0.709]	[0.001]	[0.735]	[0.657]	[0.004]	[0.006]	[0.009]	[0.003]	[0.015]
Marginal effect at $+\sigma$	4.330	3.525	-0.005	4.722	4.107	-0.031	0.207	0.289	0.433	0.617
	(0.801)	(0.487)	(0.035)	(0.876)	(0.610)	(0.038)	(0.109)	(0.127)	(0.173)	(0.218)
	[0.000]	[0.000]	[0.888]	[0.000]	[0.000]	[0.415]	[0.059]	[0.024]	[0.013]	[0.005]
p-value of nonparametric specification test	0.000	0.000	0.129	0.000	0.000	0.070	0.614	0.592	0.663	0.453
R^2	0.547	0.243	0.770	0.558	0.264	0.751	0.434	0.659	0.472	0.210
N	430	430	430	430	430	430	430	430	430	430

Notes: All columns report OLS results from estimating equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. See text and Data Appendix for more details. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table 4
Effects of Labor Demand Shocks on Housing Market
and Public Assistance Expenditures

	(1)	(2)	(3)
Dependent Variable:	Residualized Housing Values	Residualized Rental Prices	Food Stamps + Income Maintenance Expenditures
% Change in predicted employment (β)	0.718	0.838	-2.366
	(0.356)	(0.150)	(0.615)
	[0.045]	[0.000]	[0.000]
(% Change in predicted employment)^2 (δ)	-2.641	-0.998	-21.772
	(6.257)	(2.742)	(12.137)
	[0.673]	[0.716]	[0.074]
Marginal effect at $-\sigma$	0.898	0.906	-0.877
	(0.566)	(0.238)	(1.017)
	[0.114]	[0.000]	[0.390]
Marginal effect at $+\sigma$	0.537	0.770	-3.855
	(0.548)	(0.242)	(1.049)
	[0.328]	[0.002]	[0.000]
p-value of nonparametric specification test	0.328	0.606	0.149
R^2	0.142	0.097	0.401
N	430	430	430

Notes: All columns report OLS results from estimating equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts and the REIS database. The REIS database contains total county-level expenditures on Food Stamps and Income Maintenance programs. These data are aggregated to MSAs using 1990 MSA definition and adjusted per non-college capita using MSA population estimates from the Census. Final sample is a balanced panel of 215 MSAs. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. See text and Data Appendix for more details. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table 5
Effects of Alternative Measures of Labor Demand Shocks on Population

Dependent Variable: % Change in Population	(1)	(2)	(3)	(4)	(5)	(6)
% Change in predicted employment (β)	1.800	3.766	2.169	1.853	1.691	2.193
	(0.445)	(0.667)	(0.538)	(0.455)	(0.579)	(0.546)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.004]	[0.000]
(% Change in predicted employment)^2 (δ)	28.004	30.572	35.244	30.646	45.847	43.286
	(7.898)	(10.671)	(13.344)	(8.721)	(12.446)	(13.340)
	[0.000]	[0.005]	[0.009]	[0.001]	[0.000]	[0.001]
Marginal effect at $-\sigma$	-0.115	2.480	0.170	-0.214	-0.984	-0.311
	(0.838)	(0.567)	(1.097)	(0.847)	(1.105)	(1.074)
	[0.891]	[0.000]	[0.877]	[0.801]	[0.374]	[0.772]
Marginal effect at $+\sigma$	3.716	5.051	4.167	3.920	4.365	4.698
	(0.528)	(0.986)	(0.722)	(0.623)	(0.709)	(0.795)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
p-value of nonparametric specification test	0.000	0.016	0.001	0.008	0.000	0.001
R^2	0.318	0.306	0.306	0.311	0.321	0.312
N	430	430	430	430	430	430
Industries Used to Construct Predicted Employment Change						
All Industries	X					
Drop Trending Industries		X				
Drop Volatile Industries			X			
Drop Stable Industries				X		
Drop Other Industries					X	
Drop Catch-All Industries						X

Notes: All columns report OLS results from estimating equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. Column (1) reproduces the baseline specification; remaining columns construct predicted employment changes using subsets of industries. See text, Table A1, and Data Appendix for more details. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table 6
Effects of Alternative Measures of Labor Demand Shocks on Wages

Dependent Variable: % Change in Adjusted Wage	(1)	(2)	(3)	(4)	(5)	(6)
% Change in predicted employment (β)	0.520	1.180	0.387	0.531	0.478	0.687
	(0.109)	(0.208)	(0.131)	(0.102)	(0.134)	(0.126)
	[0.000]	[0.000]	[0.003]	[0.000]	[0.000]	[0.000]
(% Change in predicted employment)^2 (δ)	1.454	2.575	-0.688	1.896	3.315	0.956
	(2.422)	(4.075)	(2.761)	(2.270)	(3.417)	(2.669)
	[0.549]	[0.528]	[0.803]	[0.404]	[0.333]	[0.721]
Marginal effect at $-\sigma$	0.420	1.072	0.426	0.403	0.285	0.632
	(0.171)	(0.161)	(0.161)	(0.173)	(0.212)	(0.153)
	[0.015]	[0.000]	[0.009]	[0.021]	[0.182]	[0.000]
Marginal effect at $+\sigma$	0.619	1.288	0.348	0.659	0.671	0.742
	(0.222)	(0.345)	(0.239)	(0.194)	(0.265)	(0.237)
	[0.006]	[0.000]	[0.147]	[0.001]	[0.012]	[0.002]
p-value of nonparametric specification test	0.493	0.424	0.082	0.278	0.101	0.168
R^2	0.342	0.406	0.350	0.374	0.361	0.383
N	430	430	430	430	430	430
Industries Used to Construct Predicted Employment Change						
All Industries	X					
Drop Trending Industries		X				
Drop Volatile Industries			X			
Drop Stable Industries				X		
Drop Other Industries					X	
Drop Catch-All Industries						X

Notes: All columns report OLS results from estimating equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. Column (1) reproduces the baseline specification; remaining columns construct predicted employment changes using subsets of industries. See text, Table A1, and Data Appendix for more details. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table 7
Effects of Alternative Measures of Labor Demand Shocks on Rental Prices

Dependent Variable: % Change in Rental Prices	(1)	(2)	(3)	(4)	(5)	(6)
% Change in predicted employment (β)	0.838	1.319	0.809	0.903	0.723	0.988
	(0.150)	(0.302)	(0.172)	(0.151)	(0.175)	(0.188)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
(% Change in predicted employment)^2 (δ)	-0.998	-6.870	-2.063	0.547	0.995	-4.062
	(2.742)	(5.482)	(3.772)	(2.947)	(3.631)	(3.756)
	[0.716]	[0.212]	[0.585]	[0.853]	[0.784]	[0.281]
Marginal effect at $-\sigma$	0.906	1.608	0.926	0.866	0.665	1.223
	(0.238)	(0.247)	(0.264)	(0.240)	(0.275)	(0.241)
	[0.000]	[0.000]	[0.001]	[0.000]	[0.017]	[0.000]
Marginal effect at $+\sigma$	0.770	1.030	0.692	0.940	0.781	0.753
	(0.242)	(0.477)	(0.285)	(0.258)	(0.274)	(0.328)
	[0.002]	[0.032]	[0.016]	[0.000]	[0.005]	[0.023]
p-value of nonparametric specification test	0.606	0.447	0.412	0.581	0.064	0.494
R^2	0.097	0.116	0.076	0.109	0.063	0.110
N	430	430	430	430	430	430
Industries Used to Construct Predicted Employment Change						
All Industries	X					
Drop Trending Industries		X				
Drop Volatile Industries			X			
Drop Stable Industries				X		
Drop Other Industries					X	
Drop Catch-All Industries						X

Notes: All columns report OLS results from estimating equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. Column (1) reproduces the baseline specification; remaining columns construct predicted employment changes using subsets of industries. See text, Table A1, and Data Appendix for more details. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table 8
Alternative Sample Definitions and Alternative Specifications (Population)

	Dependent Variable: % Change in Population						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
% Change in predicted employment (β)	1.800	1.440	1.819	1.326	1.367	1.542	0.803
	(0.445)	(0.309)	(0.414)	(0.555)	(0.947)	(0.668)	(0.341)
	[0.000]	[0.000]	[0.000]	[0.018]	[0.150]	[0.022]	[0.019]
(% Change in predicted employment)^2 (δ)	28.004	14.523	24.066	25.088	32.682	39.029	18.147
	(7.898)	(3.857)	(7.525)	(7.864)	(20.336)	(23.016)	(6.673)
	[0.000]	[0.000]	[0.002]	[0.002]	[0.110]	[0.091]	[0.007]
Marginal effect at $-\sigma$	-0.172	0.417	0.124	-0.441	-0.936	-1.207	-0.475
	(0.852)	(0.369)	(0.761)	(0.901)	(1.440)	(2.114)	(0.625)
	[0.840]	[0.260]	[0.871]	[0.625]	[0.516]	[0.569]	[0.448]
Marginal effect at $+\sigma$	3.773	2.463	3.515	3.093	3.669	4.292	2.082
	(0.538)	(0.449)	(0.571)	(0.648)	(1.955)	(1.296)	(0.532)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.062]	[0.001]	[0.000]
p-value of nonparametric specification test	0.000	0.104	0.000	0.003	0.002	0.018	0.002
R^2	0.318	0.825	0.248	0.361	0.614	0.308	0.300
N	430	645	528	430	430	408	420
Alternative Samples and Alternative Specifications							
Baseline Sample	X			X	X	X	X
Add in 2000-2007		X					
Add in non-MSA regions of states			X				
Region-Specific Linear Time Trends				X			
MSA-Specific Linear Time Trends					X		
Drop Outlying 5% Shocks						X	
Compute predicted employment from CBP							X

Notes: All columns report OLS results from estimating equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. See text and Data Appendix for more details. Column (1) reproduces baseline results; remaining columns either modify baseline sample or add time-varying controls. Column (2) includes 2000-2007 changes by adding data from the American Community Survey. Column (3) includes non-MSA regions of each state as "pseudo-MSAs." Column (4) includes region-specific linear time trends. Column (5) includes MSA-specific linear time trends. Column (6) drops the 5% of observations with largest magnitude in % Change in predicted employment. Column (7) uses County Business Patterns (CBP) data to construct the predicted employment variable. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table 9
Alternative Sample Definitions and Alternative Specifications (Wages)

	Dependent Variable: % Change in Adjusted Wages						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
% Change in predicted employment (β)	0.520	0.716	0.431	0.595	0.897	0.384	0.403
	(0.109)	(0.092)	(0.096)	(0.103)	(0.245)	(0.119)	(0.086)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]
(% Change in predicted employment)^2 (δ)	1.454	-0.423	0.920	0.237	3.852	2.073	2.073
	(2.422)	(1.669)	(2.225)	(2.263)	(4.525)	(3.575)	(1.289)
	[0.549]	[0.800]	[0.679]	[0.917]	[0.396]	[0.563]	[0.109]
Marginal effect at $-\sigma$	0.417	0.745	0.366	0.578	0.626	0.238	0.257
	(0.175)	(0.124)	(0.158)	(0.165)	(0.363)	(0.267)	(0.099)
	[0.018]	[0.000]	[0.021]	[0.001]	[0.086]	[0.373]	[0.010]
Marginal effect at $+\sigma$	0.622	0.686	0.496	0.612	1.169	0.530	0.549
	(0.226)	(0.170)	(0.206)	(0.212)	(0.438)	(0.290)	(0.147)
	[0.006]	[0.000]	[0.017]	[0.004]	[0.008]	[0.069]	[0.000]
p-value of nonparametric specification test	0.493	0.025	0.317	0.595	0.191	0.322	0.346
R^2	0.342	0.087	0.337	0.406	0.547	0.340	0.344
N	430	645	528	430	430	408	420
Alternative Samples and Alternative Specifications							
Baseline Sample	X			X	X	X	X
Add in 2000-2007		X					
Add in non-MSA regions of states			X				
Region-Specific Linear Time Trends				X			
MSA-Specific Linear Time Trends					X		
Drop Outlying 5% Shocks						X	
Compute predicted employment from CBP							X

Notes: All columns report OLS results from estimating equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. See text and Data Appendix for more details. Column (1) reproduces baseline results; remaining columns either modify baseline sample or add time-varying controls. Column (2) includes 2000-2007 changes by adding data from the American Community Survey. Column (3) includes non-MSA regions of each state as "pseudo-MSAs." Column (4) includes region-specific linear time trends. Column (5) includes MSA-specific linear time trends. Column (6) drops the 5% of observations with largest magnitude in % Change in predicted employment. Column (7) uses County Business Patterns (CBP) data to construct the predicted employment variable. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table 10
Alternative Sample Definitions and Alternative Specifications (Rental Prices)

	Dependent Variable: % Change in Rental Prices						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
% Change in predicted employment (β)	0.838	0.973	0.800	0.715	0.927	0.831	0.860
	(0.150)	(0.134)	(0.134)	(0.146)	(0.363)	(0.179)	(0.121)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.011]	[0.000]	[0.000]
(% Change in predicted employment)^2 (δ)	-0.998	-0.254	0.173	-2.078	4.139	2.649	1.707
	(2.742)	(1.808)	(2.629)	(2.547)	(5.964)	(5.433)	(1.976)
	[0.716]	[0.888]	[0.948]	[0.416]	[0.488]	[0.626]	[0.389]
Marginal effect at $-\sigma$	0.908	0.991	0.788	0.861	0.636	0.644	0.740
	(0.243)	(0.161)	(0.216)	(0.222)	(0.472)	(0.380)	(0.168)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.180]	[0.092]	[0.000]
Marginal effect at $+\sigma$	0.768	0.955	0.812	0.569	1.219	1.017	0.981
	(0.246)	(0.206)	(0.241)	(0.241)	(0.628)	(0.461)	(0.200)
	[0.002]	[0.000]	[0.001]	[0.019]	[0.054]	[0.028]	[0.000]
p-value of nonparametric specification test	0.606	0.508	0.414	0.585	0.245	0.318	0.181
R^2	0.097	0.224	0.072	0.161	0.369	0.082	0.141
N	430	645	528	430	430	408	420
Alternative Samples and Alternative Specifications							
Baseline Sample	X			X	X	X	X
Add in 2000-2007		X					
Add in non-MSA regions of states			X				
Region-Specific Linear Time Trends				X			
MSA-Specific Linear Time Trends					X		
Drop Outlying 5% Shocks						X	
Compute predicted employment from CBP							X

Notes: All columns report OLS results from estimating equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. See text and Data Appendix for more details. Column (1) reproduces baseline results; remaining columns either modify baseline sample or add time-varying controls. Column (2) includes 2000-2007 changes by adding data from the American Community Survey. Column (3) includes non-MSA regions of each state as "pseudo-MSAs." Column (4) includes region-specific linear time trends. Column (5) includes MSA-specific linear time trends. Column (6) drops the 5% of observations with largest magnitude in % Change in predicted employment. Column (7) uses County Business Patterns (CBP) data to construct the predicted employment variable. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table 11
GMM Estimates of Full Model

Row	Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Housing Supply Curve		High-Skill Mobility Cost Function		Low-Skill Mobility Cost Function		H_0: $\sigma^L = \sigma^H$	Transfer Payment Elasticity	Returns to Scale	χ^2 test statistic
		σ^h	β^h	σ^H	β^H	σ^L	β^L	$\sigma^L - \sigma^H$	ψ	α	
1	Baseline Model	0.748	5.283	-0.093	-0.597	-0.070	-0.620	-0.023	-4.467	1.051	19.101
		(0.335)	(2.436)	(0.016)	(0.525)	(0.015)	(0.616)	(0.014)	(0.420)	(0.020)	[0.639]
		[0.026]	[0.031]	[0.000]	[0.257]	[0.000]	[0.314]	[0.108]	[0.000]	[0.000]	
2	No Housing; No Transfers	1.027	4.733	-0.180	-0.649	-0.289	-0.964	0.109	-4.926	1.042	25.186
		(0.314)	(2.130)	(0.017)	(0.291)	(0.022)	(0.212)	(0.016)	(0.565)	(0.019)	[0.288]
		[0.001]	[0.027]	[0.000]	[0.026]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
3	No Transfers	1.527	3.177	-0.077	-0.711	-0.137	-0.642	0.060	-5.337	1.047	20.797
		(0.338)	(1.274)	(0.015)	(0.627)	(0.017)	(0.396)	(0.014)	(0.539)	(0.017)	[0.533]
		[0.000]	[0.013]	[0.000]	[0.258]	[0.000]	[0.106]	[0.000]	[0.000]	[0.000]	
4	No Housing	1.039	4.719	-0.180	-0.667	-0.199	-0.990	0.019	-4.834	1.042	25.727
		(0.308)	(2.054)	(0.016)	(0.293)	(0.016)	(0.214)	(0.014)	(0.483)	(0.019)	[0.264]
		[0.001]	[0.022]	[0.000]	[0.024]	[0.000]	[0.000]	[0.189]	[0.000]	[0.000]	
5	$\sigma_{H,L} = 50$	1.709	3.788	-0.078	0.095	-0.056	0.185	-0.021	-4.400	1.023	29.122
		(0.505)	(1.495)	(0.013)	(0.316)	(0.013)	(0.512)	(0.005)	(0.392)	(0.025)	[0.141]
		[0.001]	[0.012]	[0.000]	[0.763]	[0.000]	[0.718]	[0.000]	[0.000]	[0.000]	
6	$\sigma_{H,L} = 0.05$	0.072	14.969	-0.108	-1.272	-0.080	-0.931	-0.028	-3.939	1.638	45.799
		(0.176)	(11.125)	(0.011)	(0.444)	(0.013)	(0.430)	(0.014)	(0.394)	(0.020)	[0.002]
		[0.684]	[0.179]	[0.000]	[0.004]	[0.000]	[0.031]	[0.050]	[0.000]	[0.000]	
7	Alternative Wage Measure	0.596	5.856	-0.026	-2.013	-0.003	-10.206	-0.023	-4.594	0.960	25.026
	(Residualized Wages)	(0.311)	(2.679)	(0.011)	(1.530)	(0.004)	(3.203)	(0.010)	(0.438)	(0.011)	[0.296]
		[0.056]	[0.029]	[0.013]	[0.189]	[0.462]	[0.002]	[0.024]	[0.000]	[0.000]	
8	Drop Labor Demand Moments	1.024	4.273	-0.100	-0.384	-0.070	-0.052	-0.029	-4.786		12.460
		(0.615)	(3.453)	(0.021)	(0.574)	(0.019)	(0.648)	(0.014)	(0.468)		[0.490]
		[0.097]	[0.217]	[0.000]	[0.504]	[0.000]	[0.936]	[0.040]	[0.000]		

Notes: All rows report estimates of the full model using a nonlinear, simultaneous equations GMM estimator. Alternate specifications are presented in each row; parameter estimates are listed in the columns. See Section 6 of main text and the Appendix for details. Asymptotic standard errors are in parenthesis and p-values are in brackets.

Table A1
Industry Categories (Top 20 List By Average National Employment Share)

Industry Name	Industry Code	Average National Empl. Share	National Employment Growth Rates Mean	Min	Max
Persistently Expanding Industries					
Eating and drinking places	641	3.99%	14.17%	1.39%	32.78%
Offices and clinics of physicians	812	0.93%	25.41%	1.28%	42.59%
Legal services	841	0.79%	32.33%	7.54%	63.76%
Computer and data processing services	732	0.75%	75.71%	20.01%	139.09%
Accounting, auditing, and bookkeeping services	890	0.47%	15.62%	4.93%	34.52%
Services incidental to transportation	432	0.45%	50.41%	1.48%	85.46%
Services to dwellings and other buildings	722	0.43%	30.51%	4.10%	73.12%
Offices and clinics of dentists	820	0.43%	26.94%	13.17%	54.00%
Personnel supply services	731	0.39%	35.46%	0.80%	62.28%
Landscape and horticultural services	20	0.34%	55.19%	5.36%	158.68%
Detective and protective services	740	0.30%	42.85%	17.52%	103.97%
Residential care facilities, without nursing	870	0.27%	86.39%	33.35%	168.29%
Drugs	181	0.27%	12.75%	3.79%	22.01%
Sporting goods, bicycles, and hobby stores	651	0.23%	17.37%	11.76%	24.11%
Veterinary services	12	0.16%	40.92%	26.99%	54.86%
Retail nurseries and garden stores	582	0.12%	61.71%	12.43%	152.46%
Museums, art galleries, and zoos	872	0.11%	56.61%	28.32%	110.27%
Offices and clinics of optometrists	822	0.05%	25.20%	18.55%	37.94%
Offices and clinics of chiropractors	821	0.05%	92.45%	14.21%	191.25%
Persistently Declining Industries					
Apparel and accessories, except knit	151	0.89%	-39.17%	-59.31%	-18.67%
Aircraft and parts	352	0.65%	-26.94%	-40.51%	-10.53%
Blast furnaces, steelworks, rolling and finishing mills	270	0.51%	-26.75%	-52.32%	-10.65%
Radio, TV, and communication equipment	341	0.48%	-35.55%	-42.74%	-27.94%
Railroads	400	0.48%	-28.70%	-51.27%	-6.72%
Yarn, thread, and fabric mills	142	0.44%	-33.50%	-55.91%	-11.45%
Newspaper publishing and printing	171	0.44%	-15.78%	-20.69%	-13.19%
Laundry, cleaning, and garment services	771	0.39%	-25.61%	-49.71%	-6.59%
Metalworking machinery	320	0.31%	-26.67%	-31.42%	-17.49%
Pulp, paper, and paperboard mills	160	0.31%	-18.36%	-32.41%	-2.25%
Motor vehicles and equipment	500	0.26%	-13.26%	-30.43%	-5.03%
Ship and boat building and repairing	360	0.26%	-21.82%	-33.81%	-6.93%
Beverage industries	120	0.20%	-15.67%	-31.90%	-1.72%
Air force	941	0.20%	-29.65%	-46.78%	-12.51%
Paperboard containers and boxes	162	0.18%	-21.47%	-33.70%	-7.28%
Variety stores	592	0.18%	-47.74%	-48.42%	-47.07%
Canned, frozen, and preserved fruits and vegetables	102	0.18%	-17.75%	-26.05%	-8.16%
Other rubber products, and plastics footwear and belting	211	0.18%	-33.14%	-61.17%	-16.64%
Navy	942	0.18%	-27.63%	-41.14%	-14.12%
Other primary metal industries	280	0.17%	-33.08%	-58.09%	-8.57%
Stable Industries					
Elementary and secondary schools	842	6.53%	3.25%	-3.59%	7.45%
All construction	60	5.94%	5.25%	1.61%	7.99%
Colleges and universities	850	2.21%	5.40%	-5.68%	16.30%
Grocery stores	601	2.06%	-0.70%	-18.49%	19.74%
Insurance	711	2.02%	-1.90%	-10.44%	2.60%
Department stores	591	1.78%	-8.54%	-19.40%	11.30%
Trucking service	410	1.54%	5.91%	1.36%	10.47%
Telephone communications	441	1.21%	-9.32%	-19.93%	3.16%
Motor vehicle dealers	612	0.95%	-1.76%	-8.87%	6.85%
Hotels and motels	762	0.95%	6.60%	-9.80%	19.21%
Groceries and related products	550	0.74%	-9.67%	-12.00%	-6.04%
Religious organizations	880	0.65%	10.21%	1.45%	19.39%
Administration of economic programs	931	0.50%	-11.06%	-15.69%	-4.20%
Beauty shops	772	0.44%	-5.35%	-16.45%	6.59%
Furniture and home furnishings stores	631	0.44%	-0.18%	-7.68%	5.36%
Sawmills, planing mills, and millwork	231	0.42%	0.28%	-13.17%	10.24%
Bus service and urban transit	401	0.40%	-8.17%	-18.13%	0.34%
Agricultural production, livestock	11	0.39%	-4.56%	-13.70%	10.29%
Public finance, taxation, and monetary policy	921	0.29%	-4.23%	-8.72%	-0.25%
Water supply and irrigation	470	0.18%	-1.98%	-4.71%	5.08%

Volatile Industries					
Justice, public order, and safety	910	2.08%	-2.02%	-42.53%	26.30%
Motor vehicles and motor vehicle equipment	351	1.31%	-9.15%	-21.15%	24.42%
National security and international affairs	932	1.02%	-14.28%	-69.36%	50.12%
Automotive repair and related services	751	0.68%	14.32%	-23.44%	30.71%
Apparel and accessory stores, except shoe	623	0.64%	-5.43%	-28.42%	30.70%
Administration of human resources programs	922	0.58%	-0.25%	-38.36%	29.24%
Management and public relations services	892	0.49%	17.01%	-31.54%	54.16%
Radio, tv, and computer stores	633	0.38%	17.95%	-32.45%	102.77%
Oil and gas extraction	42	0.38%	7.56%	-36.31%	55.97%
Computers and related equipment	322	0.37%	-5.86%	-39.46%	48.79%
Research, development, and testing services	891	0.36%	31.34%	-20.44%	106.12%
Guided missiles, space vehicles, and parts	362	0.27%	15.85%	-48.89%	62.69%
Iron and steel foundries	271	0.23%	-22.67%	-56.79%	62.60%
Scientific and controlling instruments	371	0.22%	3.56%	-22.68%	34.13%
Savings institutions, including credit unions	701	0.22%	8.96%	-32.44%	44.24%
Administration of environmental quality and housing programs	930	0.21%	1.66%	-50.06%	44.03%
Hardware, plumbing and heating supplies	521	0.20%	-2.08%	-45.76%	31.74%
Drugs, chemicals, and allied products	541	0.20%	0.38%	-26.71%	38.45%
Petroleum refining	200	0.19%	-15.09%	-32.74%	25.77%
Catalog and mail order houses	663	0.16%	11.19%	-23.37%	71.12%
Other Industries					
Hospitals	831	4.63%	5.56%	-8.14%	22.95%
Banking	700	1.76%	0.40%	-20.39%	18.42%
Real estate, including real estate-insurance offices	712	1.26%	11.18%	-14.27%	33.53%
Nursing and personal care facilities	832	1.18%	21.88%	-8.46%	74.03%
Printing, publishing, and allied industries, except newspapers	172	1.07%	-10.11%	-20.00%	10.63%
U.S. postal service	412	0.85%	-10.99%	-28.24%	1.17%
Agricultural production, crops	10	0.78%	-12.45%	-28.96%	9.86%
Engineering, architectural, and surveying services	882	0.72%	29.53%	-6.11%	67.63%
Machinery, equipment, and supplies	530	0.66%	-12.22%	-47.76%	19.83%
Child day care services	862	0.66%	39.53%	-6.98%	82.69%
Security, commodity brokerage, and investment companies	710	0.63%	29.48%	-18.73%	59.34%
Electric light and power	450	0.61%	-0.78%	-9.00%	21.04%
Air transportation	421	0.59%	-8.72%	-22.35%	13.89%
Furniture and fixtures	242	0.56%	-6.62%	-27.01%	4.43%
Drug stores	642	0.53%	4.73%	-13.18%	24.10%
Lumber and building material retailing	580	0.52%	15.50%	-4.47%	48.91%
Gasoline service stations	621	0.48%	-14.80%	-28.20%	9.14%
Fabricated structural metal products	282	0.44%	-13.01%	-27.63%	3.77%
Meat products	100	0.35%	-4.13%	-25.30%	16.68%
Radio and television broadcasting and cable	440	0.33%	21.23%	-5.68%	48.12%

Notes: All industry codes in the Census IPUMS data set are grouped into one of the five categories in this table based on employment growth during 1970-1980, 1980-1990, 1990-2000 and 2000-2007. The industries within each cataegory are then sorted based on average employment share of national population and the top 20 industries in each category are listed in this table. See Section 5 in main text for more details on the industry categories. Industries that are coded as "catch-all" industry codes are excluded from this table.

Table A2
Results for Various Measures of Public Assistance Expenditures

Dependent Variable:	(1) Food Stamps + Income Maint.	(2) Food Stamps	(3) Income Maint-enance Programs	(4) Medicare Benefits	(5) Public Medical Benefits (Medicaid)	(6) Retirement and Disability Benefits	(7) SSI Benefits	(8) UI Compen-sation	(9) Veterans Benefits	(10) Fraction of Population Disabled
% Change in predicted employment (β)	-2.366	-1.965	-4.839	-1.737	-2.150	-1.139	-1.725	-0.377	-0.914	-0.099
	(0.615)	(0.613)	(0.863)	(0.631)	(1.130)	(0.429)	(0.639)	(0.739)	(0.344)	(0.023)
	[0.000]	[0.002]	[0.000]	[0.006]	[0.058]	[0.008]	[0.008]	[0.610]	[0.008]	[0.000]
(% Change in predicted employment)^2 (δ)	-21.772	-18.731	-41.354	-22.511	-14.387	-15.743	-3.226	-31.174	1.695	0.429
	(12.137)	(11.689)	(17.538)	(12.827)	(19.519)	(8.952)	(11.779)	(11.188)	(7.369)	(0.361)
	[0.074]	[0.111]	[0.019]	[0.081]	[0.462]	[0.080]	[0.784]	[0.006]	[0.818]	[0.236]
Marginal effect at $-\sigma$	-0.877	-0.684	-2.010	-0.197	-1.166	-0.062	-1.504	1.755	-1.030	-0.128
	(1.017)	(0.997)	(1.376)	(1.184)	(1.782)	(0.830)	(1.202)	(0.961)	(0.714)	(0.029)
	[0.390]	[0.493]	[0.146]	[0.868]	[0.514]	[0.941]	[0.212]	[0.069]	[0.151]	[0.000]
Marginal effect at $+\sigma$	-3.855	-3.246	-7.667	-3.276	-3.134	-2.215	-1.945	-2.509	-0.798	-0.070
	(1.049)	(1.018)	(1.572)	(0.966)	(1.715)	(0.654)	(0.819)	(1.157)	(0.484)	(0.038)
	[0.000]	[0.002]	[0.000]	[0.001]	[0.069]	[0.001]	[0.018]	[0.031]	[0.101]	[0.070]
p-value of nonparametric specification test	0.149	0.230	0.008	0.144	0.026	0.057	0.418	0.015	0.437	0.081
R^2	0.401	0.435	0.273	0.796	0.695	0.556	0.531	0.071	0.339	0.045
N	430	430	430	430	430	430	430	430	430	430

Notes: All columns report OLS results from estimating equation (7). Data for dependent variables come from the REIS, except for column (10) which uses Census data on disability in the adult population. Final sample is a balanced panel of 215 MSAs. Dependent variable is always the percentage change across periods except for column (10) which reports percentage point changes. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. See Section 4 in main text and Data Appendix for more details. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table A3
Robustness Dropping Each Region

	Dependent Variable: % Change in Population									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	All Regions	Drop New England	Drop Middle Atlantic	Drop East North Central	Drop West North Central	Drop South Atlantic	Drop East South Central	Drop West South Central	Drop Mountain	Drop Pacific
% Change in predicted employment (β)	1.800	1.657	2.141	1.658	1.766	1.705	1.836	1.795	1.868	1.762
	(0.445)	(0.447)	(0.545)	(0.539)	(0.471)	(0.380)	(0.451)	(0.468)	(0.455)	(0.504)
	[0.000]	[0.000]	[0.000]	[0.002]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]
(% Change in predicted employment)^2 (δ)	28.004	27.833	22.633	29.574	29.336	20.861	27.263	28.835	31.104	30.666
	(7.898)	(7.975)	(8.980)	(9.265)	(8.347)	(7.494)	(8.002)	(8.161)	(8.555)	(8.181)
	[0.000]	[0.001]	[0.013]	[0.002]	[0.001]	[0.006]	[0.001]	[0.001]	[0.000]	[0.000]
Marginal effect at -σ	-0.115	-0.246	0.593	-0.365	-0.240	0.279	-0.029	-0.177	-0.259	-0.335
	(0.838)	(0.869)	(1.032)	(1.021)	(0.898)	(0.656)	(0.854)	(0.873)	(0.865)	(0.883)
	[0.891]	[0.777]	[0.566]	[0.721]	[0.789]	[0.671]	[0.973]	[0.840]	[0.765]	[0.705]
Marginal effect at +σ	3.716	3.561	3.688	3.680	3.772	3.132	3.700	3.767	3.995	3.859
	(0.528)	(0.490)	(0.532)	(0.584)	(0.538)	(0.620)	(0.526)	(0.545)	(0.593)	(0.596)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
p-value of nonparametric specification test	0.000	0.000	0.042	0.000	0.000	0.082	0.000	0.000	0.000	0.000
R^2	0.318	0.324	0.338	0.335	0.326	0.229	0.319	0.326	0.300	0.306
N	430	408	382	344	402	360	406	372	404	366

Notes: All columns report OLS results from estimating equation (7). Final sample is a balanced panel of 215 MSAs. Columns report results from dropping one of the nine Census regions. Dependent variable is always the percentage change in population across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. See Section 4 in main text and Data Appendix for more details. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table A4
Alternative Specifications of Adjustment Dynamics (Population)

Dependent Variable: % Change in Population	(1)	(2)	(3)	(4)	(5)
% Change in predicted employment (t)	1.800	0.783			1.260
	(0.445)	(0.342)			(1.373)
	[0.000]	[0.023]			[0.360]
(% Change in predicted employment)^2 (t)	28.004	18.224			-7.186
	(7.898)	(6.699)			(19.906)
	[0.000]	[0.007]			[0.718]
% Change in predicted employment $(t-1)$			0.689		-0.920
			(0.336)		(1.767)
			[0.041]		[0.603]
(% Change in predicted employment)^2 $(t-1)$			19.837		18.955
			(7.582)		(29.947)
			[0.010]		[0.527]
% Change in predicted employment $(t-2)$				0.476	0.333
				(0.284)	(0.692)
				[0.096]	[0.631]
(% Change in predicted employment)^2 $(t-2)$				10.668	5.189
				(3.435)	(6.500)
				[0.002]	[0.426]
Marginal effect at $-\sigma$ **(A)**	-0.115	-0.491	-0.699	-0.270	-0.537
	(0.838)	(0.630)	(0.713)	(0.459)	(0.754)
	[0.891]	[0.436]	[0.329]	[0.557]	[0.477]
Marginal effect at $+\sigma$ **(B)**	3.716	2.058	2.076	1.222	1.883
	(0.528)	(0.525)	(0.528)	(0.258)	(0.579)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]
p-value of test **(A) = (B)**	0.000	0.007	0.010	0.002	0.033
p-value of nonparametric specification test	0.000	0.002	0.025	0.439	
R^2	0.318	0.300	0.306	0.304	0.310
N	430	420	420	420	420
Alternative Samples and Alternative Specifications					
Compute predicted employment from Census	X				
Compute predicted employment from CBP		X	X	X	X

Notes: All columns report OLS results from estimating variants of equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs in column (1) and 210 MSAs in remaining columns. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. See text and Data Appendix for more details. Column (1) reproduces baseline results, while remaining columns construct predicted employment by using the County Business Patterns (CBP) instead of the Census. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Table A5
Alternative Specifications of Adjustment Dynamics (Wages)

Dependent Variable: % Change in Adjusted Wages	(1)	(2)	(3)	(4)	(5)
% Change in predicted employment (t)	0.522	0.402			0.838
	(0.109)	(0.086)			(0.378)
	[0.000]	[0.000]			[0.028]
(% Change in predicted employment)^2 (t)	1.458	2.109			1.603
	(2.427)	(1.291)			(3.495)
	[0.549]	[0.104]			[0.647]
% Change in predicted employment ($t-1$)			0.347		-0.284
			(0.082)		(0.382)
			[0.000]		[0.458]
(% Change in predicted employment)^2 ($t-1$)			1.832		0.398
			(1.196)		(3.133)
			[0.127]		[0.899]
% Change in predicted employment ($t-2$)				0.230	-0.134
				(0.069)	(0.194)
				[0.001]	[0.489]
(% Change in predicted employment)^2 ($t-2$)				0.591	0.183
				(0.729)	(1.336)
				[0.419]	[0.891]
Marginal effect at $-\sigma$ **(A)**	0.420	0.255	0.219	0.189	0.264
	(0.174)	(0.099)	(0.103)	(0.085)	(0.100)
	[0.017]	[0.011]	[0.034]	[0.028]	[0.009]
Marginal effect at $+\sigma$ **(B)**	0.624	0.550	0.476	0.272	0.575
	(0.226)	(0.146)	(0.130)	(0.086)	(0.164)
	[0.006]	[0.000]	[0.000]	[0.002]	[0.001]
p-value of test **(A)** = **(B)**	0.549	0.104	0.127	0.419	0.123
p-value of nonparametric specification test	0.494	0.348	0.002	0.013	
R^2	0.342	0.344	0.337	0.327	0.348
N	431	420	420	420	420
Alternative Samples and Alternative Specifications					
Compute predicted employment from Census	X				
Compute predicted employment from CBP		X	X	X	X

Notes: All columns report OLS results from estimating variants of equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs in column (1) and 210 MSAs in remaining columns. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. See text and Data Appendix for more details. Column (1) reproduces baseline results, while remaining columns construct predicted employment by using the County Business Patterns (CBP) instead of the Census. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

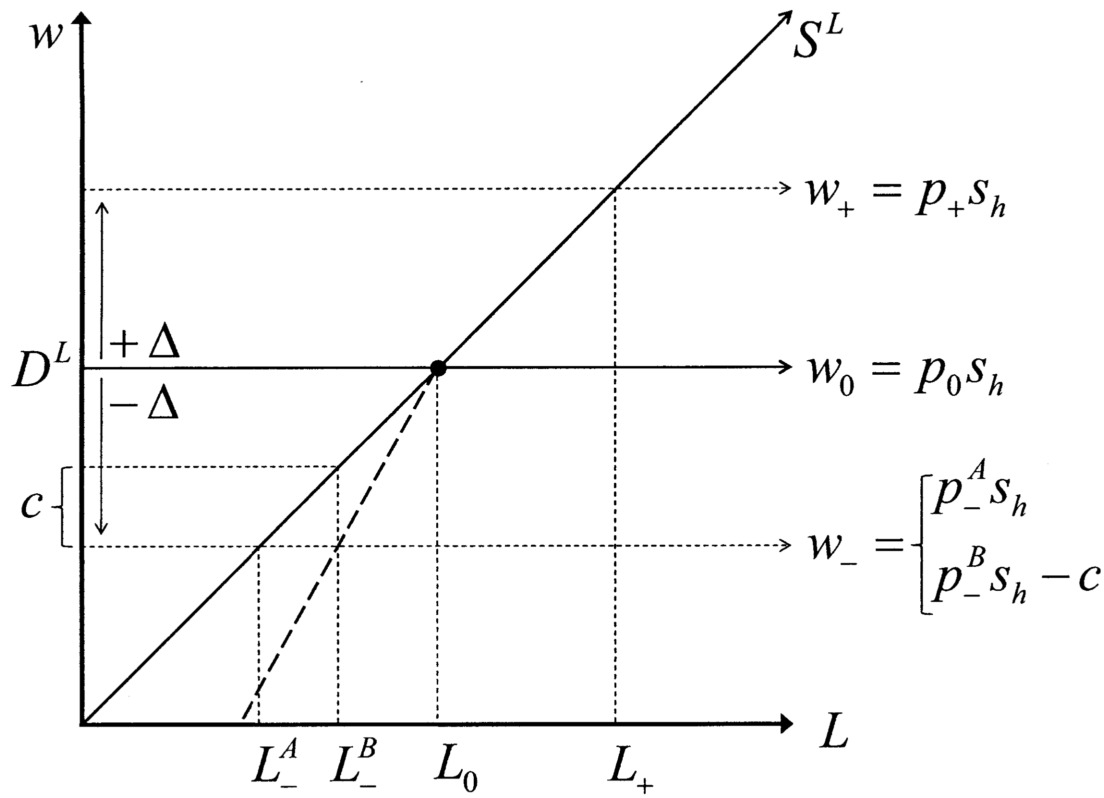
Table A6

Alternative Specifications of Adjustment Dynamics (Rental Prices)

	Dependent Variable: % Change in Rental Prices				
	(1)	(2)	(3)	(4)	(5)
% Change in predicted employment (t)	0.836	0.865			1.466
	(0.151)	(0.121)			(0.556)
	[0.000]	[0.000]			[0.009]
(% Change in predicted employment)^2 (t)	-1.001	1.753			-0.430
	(2.756)	(1.990)			(5.465)
	[0.717]	[0.380]			[0.937]
% Change in predicted employment ($t-1$)			0.773		-0.147
			(0.122)		(0.589)
			[0.000]		[0.803]
(% Change in predicted employment)^2 ($t-1$)			1.587		1.592
			(1.893)		(5.638)
			[0.403]		[0.778]
% Change in predicted employment ($t-2$)				0.518	-0.406
				(0.112)	(0.282)
				[0.000]	[0.151]
(% Change in predicted employment)^2 ($t-2$)				0.150	0.552
				(1.214)	(2.240)
				[0.902]	[0.806]
Marginal effect at $-\sigma$ **(A)**	0.906	0.742	0.662	0.508	0.791
	(0.243)	(0.169)	(0.164)	(0.119)	(0.177)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Marginal effect at $+\sigma$ **(B)**	0.766	0.987	0.884	0.529	1.036
	(0.246)	(0.199)	(0.196)	(0.158)	(0.219)
	[0.002]	[0.000]	[0.000]	[0.001]	[0.000]
p-value of test **(A)** = **(B)**	0.717	0.380	0.403	0.902	0.436
p-value of nonparametric specification test	0.604	0.178	0.184	0.013	
R^2	0.097	0.141	0.124	0.090	0.152
N	431	420	420	420	420
Alternative Samples and Alternative Specifications					
Compute predicted employment from Census	X				
Compute predicted employment from CBP		X	X	X	X

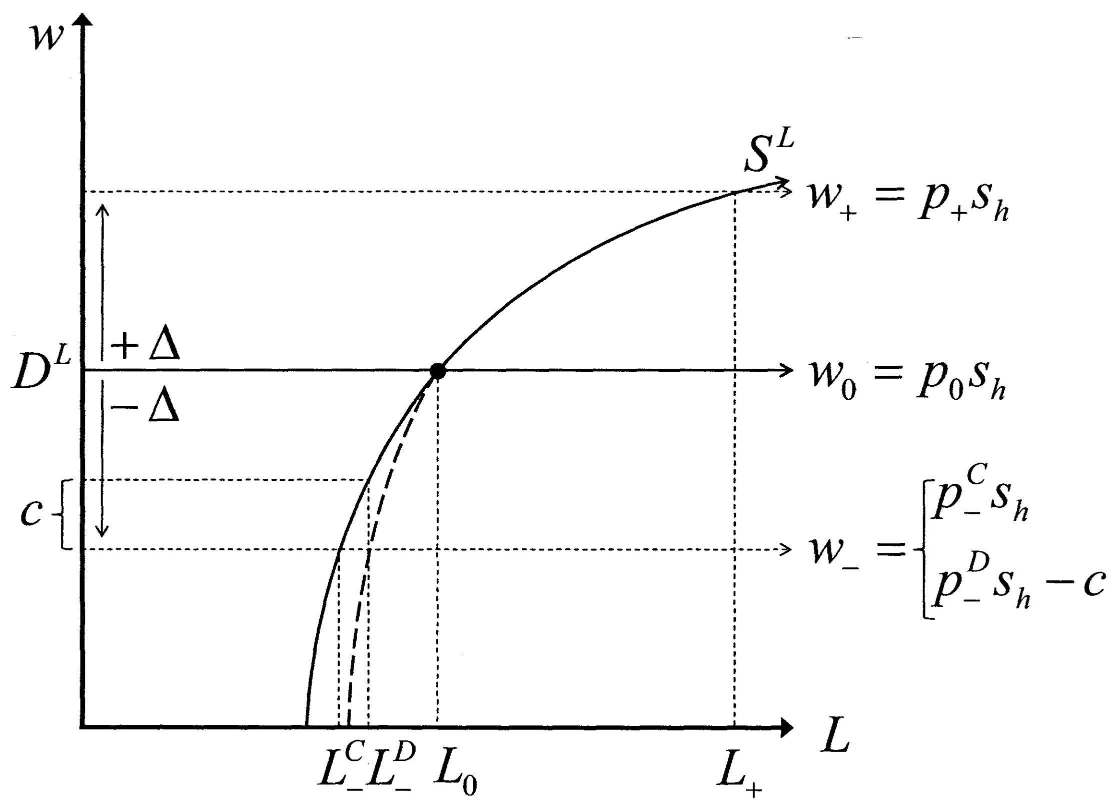
Notes: All columns report OLS results from estimating variants of equation (7). Data come from IPUMS 1980, 1990, and 2000 census extracts. Final sample is a balanced panel of 215 MSAs in column (1) and 210 MSAs in remaining columns. Dependent variable is always the percentage change across periods. The % Change in predicted employment is formed by interacting cross-sectional differences in industrial composition with national changes in industry employment shares. See text and Data Appendix for more details. Column (1) reproduces baseline results, while remaining columns construct predicted employment by using the County Business Patterns (CBP) instead of the Census. All specifications include year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parenthesis and p-values are in brackets.

Figure 1: Constant Housing Supply Elasticity



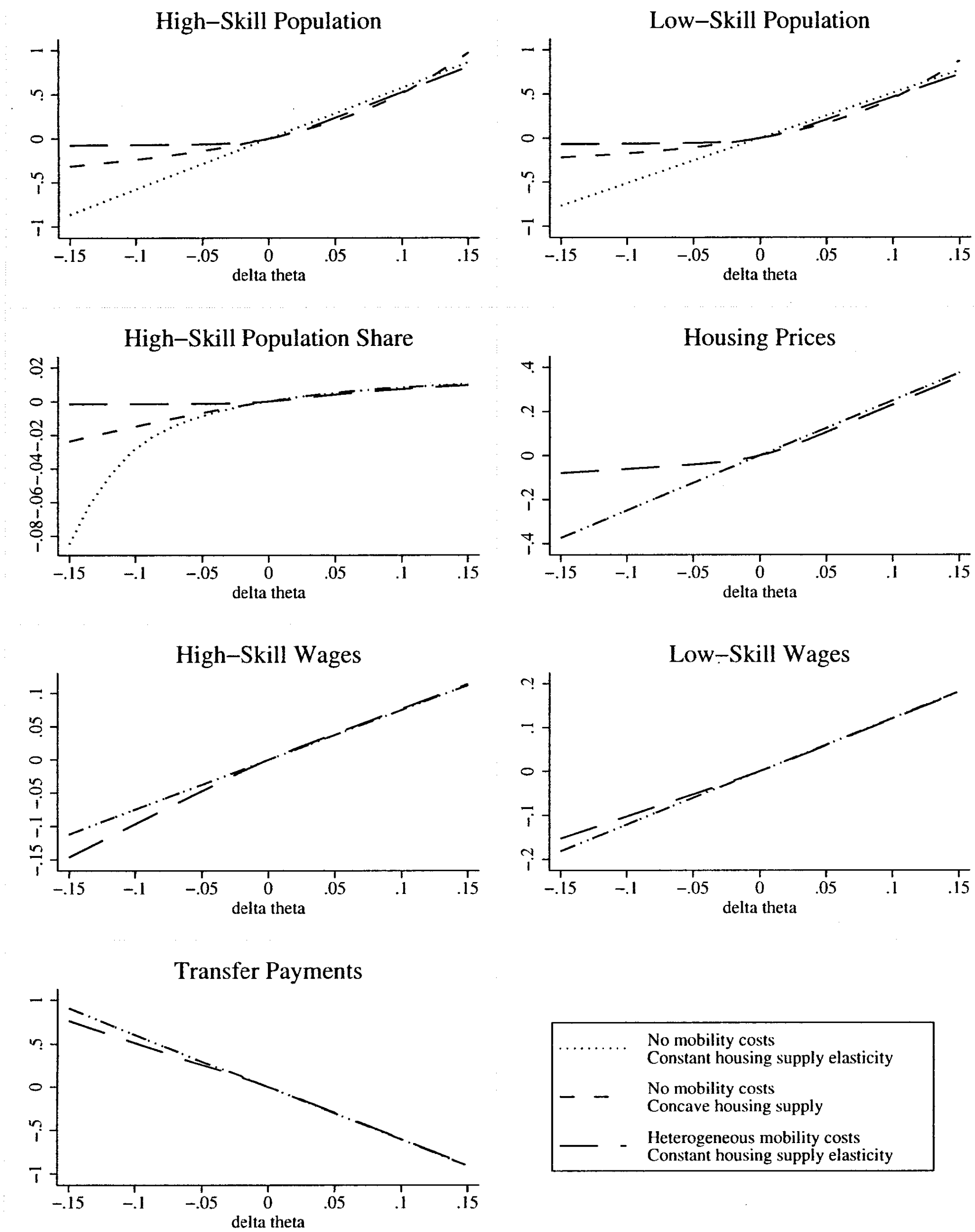
This figure displays the equilibrium response when the housing supply elasticity is constant. The initial equilibrium wages, labor supply, and housing prices are given by the dot in the center of the figure. An exogenous increase in wages encourages in-migration until labor supply rises to L_+. At this point, housing prices have risen to completely offset the increase in wages, restoring the no-arbitrage condition for workers. If there are no mobility costs, then the equilibrium response of an equal-sized exogenous decrease in wages is *symmetric*, as shown by L_-^A. If out-migration is costly, however, then following a negative shock, the marginal out-migrant must be indifferent between staying and paying c to out-migrate. These mobility costs cause both population and housing prices to respond *asymmetrically*: positive shocks increase population and housing prices more than negative shocks reduce them.

Figure 2: Concave Housing Supply Curve



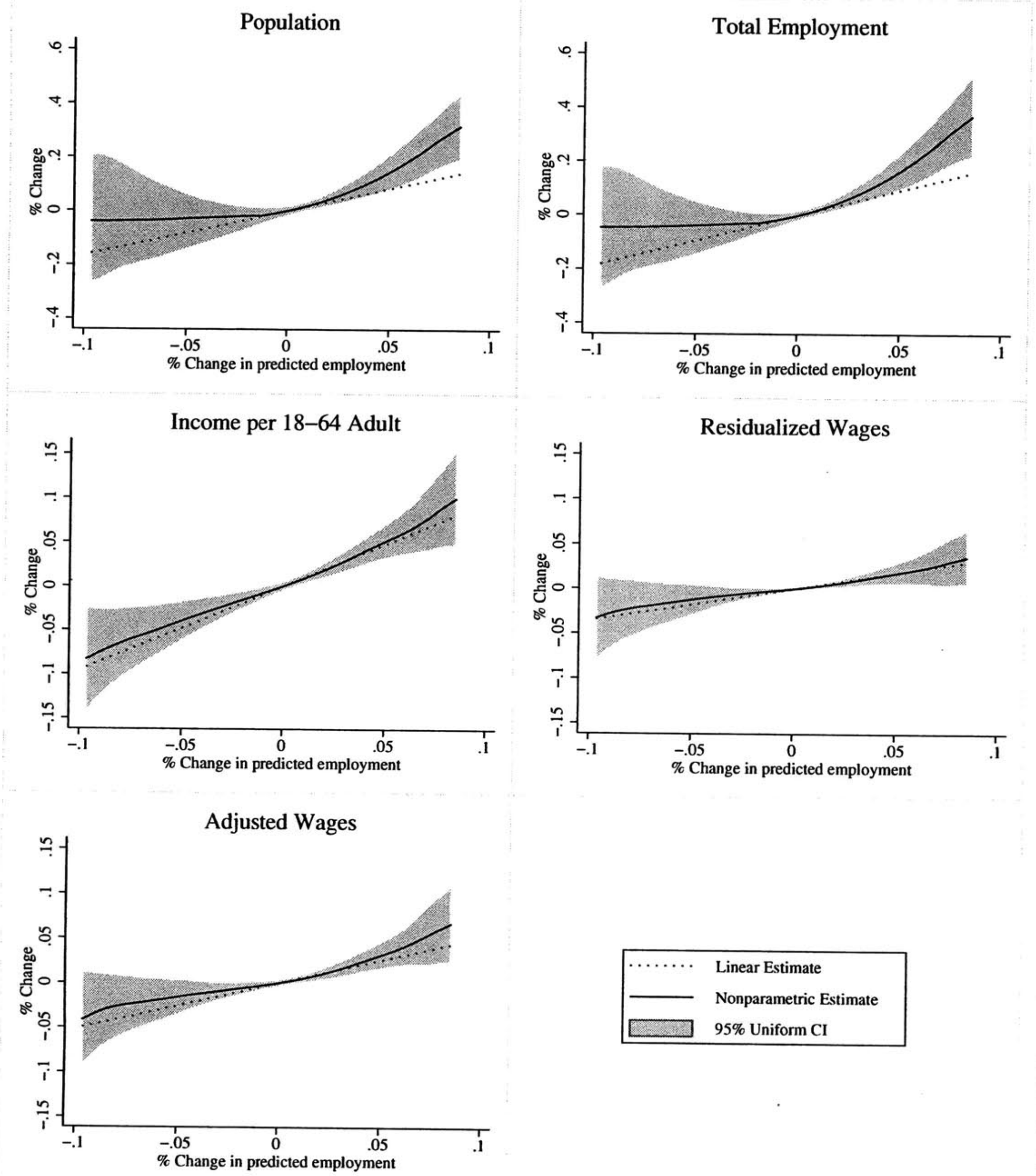
This figure displays the equilibrium response when the housing supply curve is concave. As the main text and Appendix describe in more detail, a concave housing supply curve is consistent with a durable housing stock that is not destroyed once created. As in figure 1, the initial equilibrium wages, labor supply, and housing prices are given by the dot in the center of the figure. An exogenous increase in wages encourages in-migration until labor supply rises to L_+. At this point, housing prices have risen to completely offset the increase in wages, restoring the no-arbitrage condition for workers. If there are no mobility costs, then housing prices still respond symmetrically (p_-^C). Intuitively, housing costs still must adjust to exactly offset the wage changes. Only population responds asymmetrically (as shown by L_-^C). If workers have mobility costs, then the asymmetry of the population response is even greater (see L_-^D), and in this case housing prices also respond asymmetrically.

Figure 3: Model Simulations



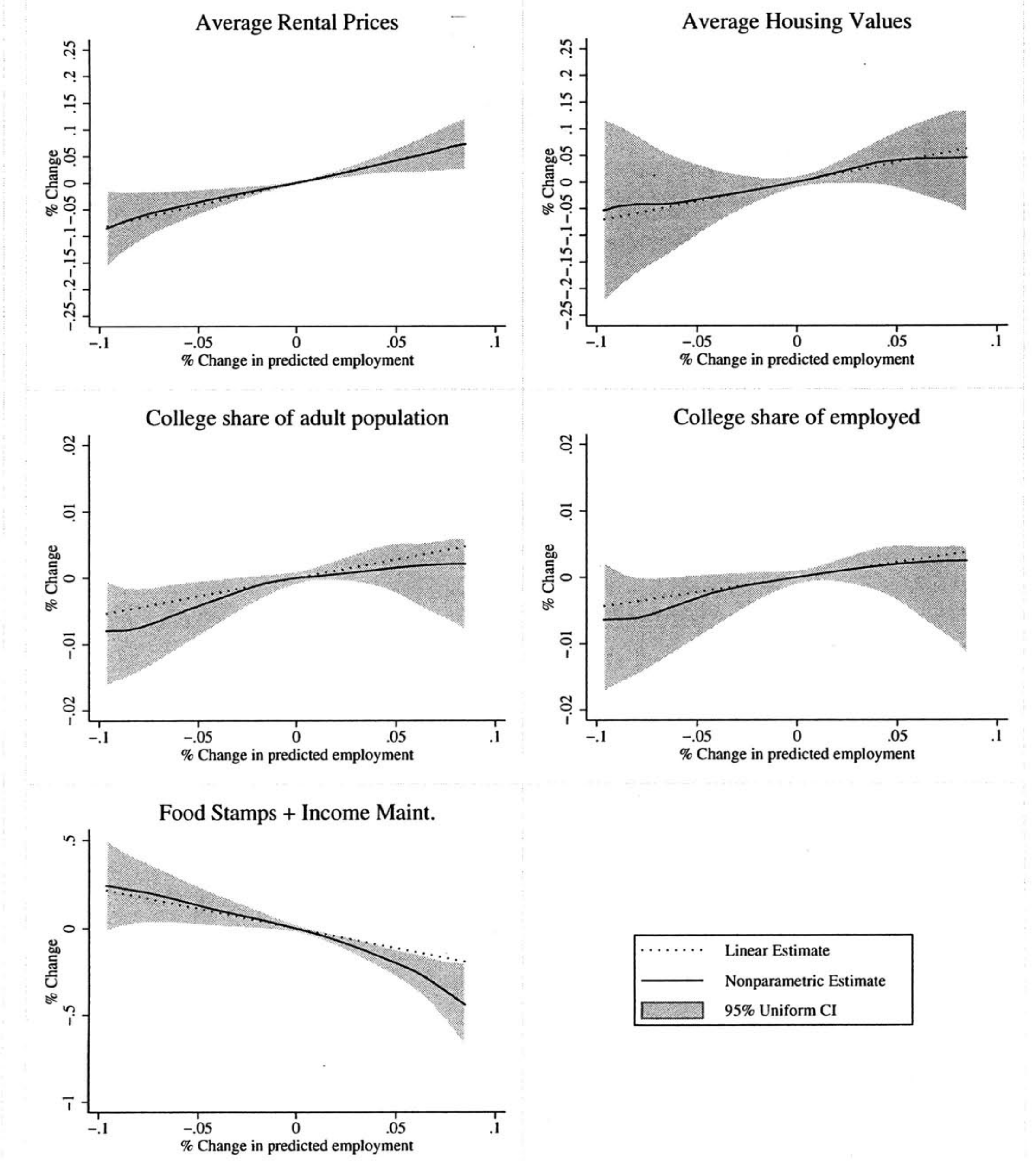
This figure displays simulated data from the model described in Section 2. See the Appendix for more details on the simulation. The graphs clarify that an asymmetric response of population to the labor demand shock (delta theta) indicates the existence of a concave housing supply curve and/or the existence of heterogeneous mobility costs. The response of housing prices isolates the importance of mobility costs.

Figure 4: Reduced Form Results



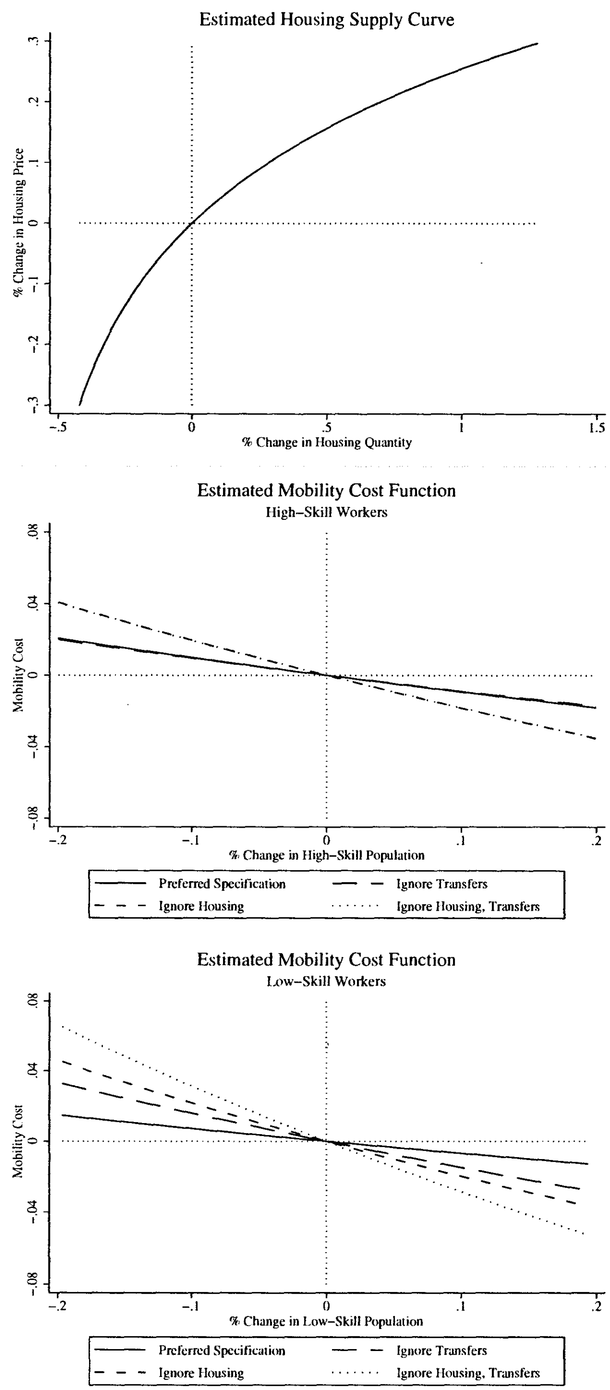
This figure reports nonparametric reduced form estimates using U.S. Census data and REIS data. See Appendix for details on the data set. All graphs are nonparamaetric local linear regressions. All results include year fixed effects in the nonparametric model. The estimates are constrained to be monotonic following the rearrangement procedure of Chernozhukov, Fernandez-Val, and Galichon (2003). The 95 percent uniform confidence intervals are computed using 10,000 bootstrap replications, resampling MSAs with replacement. In each bootstrap step, an undersmoothed local linear bandwidth is chosen following Hall (1992).

Figure 5: Reduced Form Results, Continued



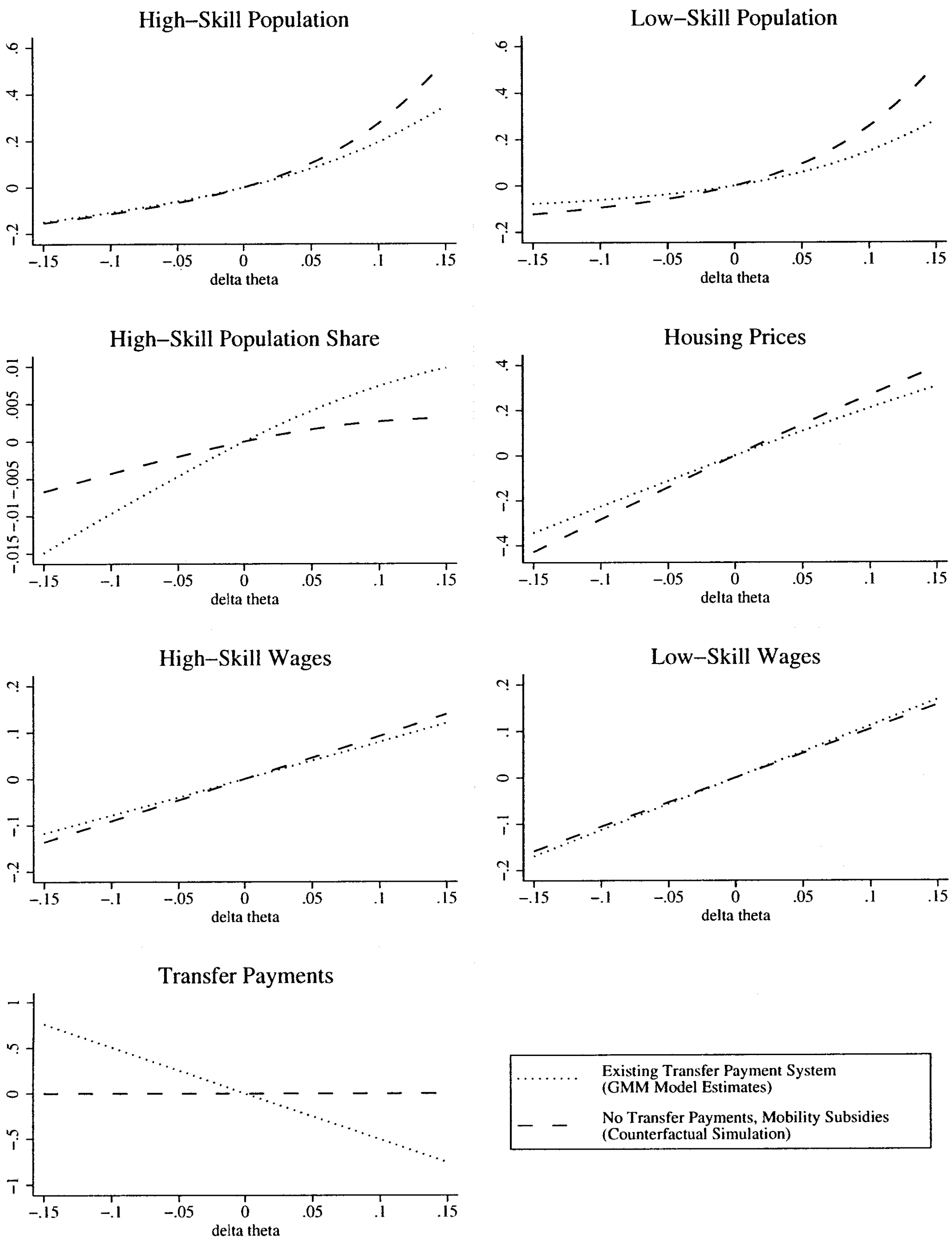
This figure reports nonparametric reduced form estimates using U.S. Census data and REIS data. See Appendix for details on the data set. All graphs are nonparamaetric local linear regressions. All results include year fixed effects in the nonparametric model. The estimates are constrained to be monotonic following the rearrangement procedure of Chernozhukov, Fernandez-Val, and Galichon (2003). The 95 percent uniform confidence intervals are computed using 10,000 bootstrap replications, resampling MSAs with replacement. In each bootstrap step, an undersmoothed local linear bandwidth is chosen following Hall (1992).

Figure 6: GMM Estimates



This figure reports GMM estimates of the full model. The top figure presents the housing supply curve that is estimated in the baseline model (Table 11, row 1). The middle and bottom figures report estimated mobility functions under various assumptions about housing expenditure shares and transfer payments. See Section 6 and the Appendix for more details on the GMM estimation.

Figure 7: Counterfactual Simulation



This figure reports simulations based on GMM estimates of the full model. The GMM estimates are used to run simulations similar to those presented in Figure 3. The graphs report results of two simulations: (1) simulation based on estimates of the baseline GMM model using the existing transfer payment system and (2) counterfactual simulation based on same estimates but transfer payment system is replaced with mobility subsidies which reduce mobility costs by 50%.

Chapter 2

Income and Health Spending: Evidence from Oil Price Shocks[1]

(Joint work with Amy Finkelstein, MIT and Daron Acemoglu, MIT)

The dramatic rise in health care expenditures is one of the notable economic trends of the postwar era. As seen in Figure 1, health care expenditure as a share of GDP in the United States has more than tripled over the last half century, from 5 percent in 1960 to 16 percent in 2005 (CMS, 2006). A common conjecture is that the rise in the share of income spent on health care expenditures is a direct, or at least a natural, consequence of the secular increase in living standards—because health care is a "luxury good".[2] The *Economist* magazine stated this as a "conventional wisdom" in 1993, writing:

> "As with luxury goods, health spending tends to rise disproportionately as countries become richer." (quoted in Blomqvist and Carter, 1997, p. 27).

[1]We are grateful to Amitabh Chandra, Bob Hall, Guy Michaels, Joe Newhouse, and participants at the NBER Health Care meetings and NBER Labor Studies Summer Institute for helpful comments, to James Wang for outstanding research assistance, and to the National Institute of Aging (NIA grant numbers P30-AG012810 and T32-AG000186) for financial support.

[2]Throughout we use the term "luxury good" to designate an empirical income elasticity greater than one (and similarly "necessity" refers to an elasticity less than one). This responsiveness to income may result from preferences, policy or other factors.

This view has recently been forcefully articulated by Hall and Jones (2007). They argue that the optimal share of spending on health increases as incomes rise, since spending money on life extension allows individuals to escape diminishing marginal utility of consumption within a period. The Hall-Jones view also receives indirect support from the very high estimates of the value of life and value of health provided by Nordhaus (2003) and Murphy and Topel (2003, 2006). The fact that most other OECD countries have also experienced substantial growth in their health sector over the last half century (OECD, 2004) also makes the secular rise in incomes a natural candidate to explain the rise in the health share of GDP in the United States.

Understanding the extent to which the rise in the health share of GDP is a direct consequence of the rise in living standards is important for several reasons. First, it enables a proper accounting of the notable growth in the US (and OECD) health care sector over the last half century. Second, it is necessary for forecasting how health care spending is likely to evolve in coming years. Finally, it is a crucial first step towards an assessment of the optimality of the growth of the health care sector. In particular, if health spending is strongly increasing in income, so that rising income can explain most or all of the rising health share, it would be more likely that the increasing share of GDP allocated to health is socially optimal.[3]

The relationship between income and health spending is the subject of a voluminous empirical literature. Remarkably, however, virtually all existing estimates are based on simple correlations of income and health care spending, across individuals, across countries, or over time. These correlations are consistent with income elasticities ranging from close to zero to substantially above one.[4] In light of the paucity of existing evidence, Hall and Jones (2007) conclude their paper by stating that "Our

[3]Of course a large role for income would only be suggestive, not dispositive. A systematic analysis of social optimality would also have to consider potential externalities in health provision and in health R&D, as well as informational and institutional constraints in the health care market.

[4]OECD (2006) provides a recent survey of the large empirical literature on the correlation between income and health spending (see particularly Annex 2B). The cross-sectional relationship across individuals between income and health spending tends to be small or negative (e.g., Newhouse and Phelps 1976). In contrast, cross-country analysis tends to suggest income elasticities greater than 1 (e.g., Newhouse 1977, Gerdtham and Jonsson 2000), as do time-series analyses of the relationship between income growth and growth in health spending for individual countries (e.g., Fogel 1999).

model makes the strong prediction that if one looks hard enough and carefully enough, one ought to be able to see income effects [with elasticities above 1] in the micro data. Future empirical work will be needed to judge this prediction."

Our objective is to provide "causal" estimates of the effect of income on aggregate health spending. There are (at least) two important challenges in this exercise. The first is that income and health covary at the individual or regional level for a variety of reasons. Therefore, simple correlations are unlikely to reveal the causal effect of income on health spending.

A second challenge is that an investigation of the role that rising income plays in the growth of the health care sector requires incorporating the general equilibrium effects of income on health spending. Partial and general equilibrium income elasticities may differ for a variety of reasons. For example, the general equilibrium effect of rising income may be larger than the partial equilibrium effect if an increase in the demand for health care from a community (a "general equilibrium change") prompts changes in medical practices, including the adoption (and possibly development) of new technologies.[5] Alternatively, if the supply of health care is less than perfectly elastic and the price elasticity of demand for health care is greater than one, the responsiveness of health care expenditures to an increase in income may be lower in general equilibrium than in partial equilibrium. In addition, changes in income may also affect health care policy through a variety of political economy channels, either magnifying or curtailing the direct effect of income on health expenditures. Many of the potential general equilibrium effects are "local" in the sense that they result from changes in incomes in a particular region or local economy. These effects can be detected by looking at the response of health spending to income in the local economy. In addition, there may also exist national or even global general equilibrium effects, which will be harder to detect empirically.

We confront both of these challenges. By exploiting potentially exogenous variation in local area incomes, we attempt to estimate causal elasticities that incorporate

[5]Finkelstein (2007), for example, argues that, for such reasons, the general equilibrium effect of health insurance coverage on health spending is larger than the partial equilibrium effect.

local general equilibrium effects. On the basis of our estimates and additional evidence, we also argue below that national or global general equilibrium income effects are unlikely to be significant in this instance.

Our strategy is to exploit the time-series variation in global oil prices between 1970 and 1990, which impacted incomes differentially across different parts of the (Southern) United States that vary in the oil intensity of the local economy. In our baseline specification we approximate local economies by economics sub-regions (ESRs), which consist of groups of counties within a state that have strong economic ties. We focus on the South of the United States to increase the comparability of the ESRs, in particular to minimize the likelihood of differential trends in health care expenditure driven by other factors. Our empirical strategy exploits the interaction between global oil prices and ESR-level importance of oil in the economy as an instrument for income. Our main proxy for the importance of oil is the size of pre-existing oil reserves in an ESR. The identifying assumption is that the interaction between global oil price changes and local oil reserves should have no effect on changes in the demand for health care, except through income. We provide several pieces of evidence that are supportive of the validity of this identifying assumption. Using this instrumental-variable strategy we estimate an elasticity of ESR-level hospital spending with respect to ESR-level income of 0.72 (standard error = 0.21). Point estimates of the income elasticity from a wide range of alternative specifications fall on both sides of our baseline estimate, but are almost always less than 1.

Because our instrument impacts incomes at the ESR level (rather than individual income), our estimates correspond to local general equilibrium effects of income changes, but will not capture any global or national general equilibrium effects.[6] Of particular concern is that if the growth of the health care market resulting from the rise in global incomes induced more innovation, our estimates may not incorporate the implications of these induced innovations on health expenditures. Our analysis suggests that significantly larger elasticities resulting from these induced innovation

[6]We also present results at the state rather than ESR level. This reduces our cross-sectional variation in oil intensity but alllows us to capture general equilibrium effects at a higher level of geographic aggregation than the ESR. The results are similar.

general equilibrium effects are unlikely for two reasons. First, the same induced innovation effects working at the national or global level should manifest themselves as increased technology adoption or entry of new hospitals at the local (ESR) level. However, we find no statistically or substantively significant effects of local income on hospital entry or on various measures of technology adoption at the ESR level. In this light, a significant *global* induced innovation effect seems unlikely. Second, technological change should be more rapid for sectors that are expanding faster than others (e.g., Acemoglu, 2002, Acemoglu and Linn, 2004). Since health care appears to have an income elasticity less than one, induced innovations should relatively favor the non-health sectors that have an income elasticity above one.[7]

We therefore use our local general equilibrium income elasticity estimate to perform a back of the envelope calculation of the role that rising income has played in the rising U.S. health share. Our central point estimate of 0.72 suggests that rising income would be associated with a modest decline in the health share of GDP. Perhaps more informatively, the upper end of the 95 percent confidence interval of this estimate is 1.13; this allows us to reject the hypothesis that rising real income explains more than 0.5 percentage points of the 11 percentage point increase in the health share of US GDP between 1960 and 2005.

We explore below several potentially important caveats to this out of sample extrapolation. In particular, our empirical work focuses primarily on hospital expenditures from the American Hospital Association data (rather than on total health expenditures). Hospital spending is the single largest component of total health care spending, and the time-series evidence in Figure 1 suggests that hospital and

[7]The fact that income induced innovations in the health care sector are not likely to be quantiatively important does not imply that induced innovation more generally is not important in the health care sector. Indeed, there is substantial empirical evidence of induced innovation effects in the health care sector arising through such mechanisms such as expected market size (Acemoglu and Linn, 2004, Finkelstein, 2004, Finkelstein, 2007) and relative factor prices (Acemoglu and Finkelstein, 2008). Moreover, the evidence in Acemoglu and Linn (2004) suggests that whether or not rising income produces induced innovation effects in the health care sector will depend on whether the heath sector expands relative to other sectors in response to rising income. Given that our estimates suggest that rising incomes increase the relative market sizes of other sectors more than that of health care, this suggests that induced innovations arising from increased income are more likely to be directed towards these other sectors rather than health care.

non-hospital components of health care have grown proportionally over the last half century. If income elasticities were substantially higher for the non-hospital components of health expenditures, and if the rise in income over this time period were the major driver of the increase in health expenditures, we should (all else equal) see a decline in the hospital share of total health expenditures. This suggests that income elasticities of hospital and non-hospital components of health expenditures should be similar. We also draw on additional data sources to provide suggestive empirical evidence that the income elasticities of hospital expenditures and overall health expenditures are similar. This evidence bolsters our belief that our elasticity estimates for hospital spending are likely to be representative of those for total health expenditures.

A final point that warrants emphasis at the outset is that our empirical strategy estimates the effect of rising incomes on health care spending in the recent US context. This empirical relationship is undoubtedly partly shaped by several specific institutional features of the US health care system. Our evidence does not therefore directly address the question of whether health care is a "luxury good" in households' utility function as hypothesized by Hall and Jones (2007).

To our knowledge, our paper represents the first empirical attempt to estimate the causal general equilibrium income elasticity of health spending.[8] Indeed, we are only aware of two prior studies that attempt to estimate the "causal effect" of income on health spending; both estimate the partial equilibrium effect of income on own health spending. Moran and Simon (2006) use the Social Security notch cohort to examine the effect of plausibly exogenous variation in an elderly individual's income

[8]Our empirical strategy is related to that used by Michaels (2007) to estimate the long-run consequences of resource-based specialization, and to those in Buckley (2003) and Black, McKinnish and Sanders (2005). Michaels also exploits variation in oil abundance across county groups within the US South and studies the consequences of the availability of greater oil resources on changes in the sectoral composition of employment and in education. Buckley (2003) exploited the same source of variation within Texas to investigate the effect of income on marriage and divorce. Black, McKinnish and Sanders (2005) use a similar strategy focusing the coal boom and bust. Davis and Haltiwanger (2001) investigate the impact of changes in oil prices on sectoral job creation and job destruction. Kilian (2008) provides a review of the literature on the economic effects of energy price shocks on a variety of different sectors and macroeconomic aggregates. None of these papers study the effect of income on the health care sector.

on the elderly's prescription drug use; they estimate an elasticity of drug use with respect to income of above one. The Rand Health Insurance Experiment finds that a small, unanticipated, temporary increase in own income has no significant impact on own health expenditures or utilization (Newhouse et al., 1993, p. 78).[9]

The rest of the paper proceeds as follows. Section 2.1 describes our empirical strategy and data. Section 2.2 contains our main results. It shows the first-stage relationship between ESR income and our instrument, presents our instrumental variable estimates of the local general equilibrium income elasticity of hospital expenditures and their components, and investigates whether national and global general equilibrium income effects are likely to be quantitatively important. Section 2.3 discusses the implications of our elasticity estimate for the role of rising income in explaining the rise in the health share of GDP in the United States; it also discusses in some depth some of the most salient potential threats to extrapolating from our estimates in this manner, and summarizes our robustness analysis. Section 2.4 concludes. The on-line appendices (which are not for publication) provide further theoretical and empirical results.

2.1 Empirical Strategy and Data

2.1.1 Motivating Theory

We start with a simple theoretical model which provides a framework for interpreting our results. The model clarifies the distinction between three different income elasticities: the "partial equilibrium" income elasticity that measures the responsiveness of health spending to a change in an individual's income, the "local general equilibrium" income elasticity from a change in an area's income, and the "global general equilibrium" income elasticity that measures the responsiveness of health spending

[9]These results are from the so-called Super Participation Incentive in which a sub-sample of families were given an unanticipated, small (a maximum of $250 in the mid 1970s) additional lump sum payment for one year in the penultimate year of the experiment. Note that this sub-experiment was not designed to estimate the income elasticity of demand for health care but rather to test whether the income side payments made to families as part of the experimental design (whose focus was to estimate the effect of cost sharing) impacted utilization.

to national or global income changes.

Consider an individual i residing in area j at time t with a utility function given by

$$\pi\left(Q_{jt}h_{ijt}\right)u\left(c_{ijt}\right), \tag{2.1}$$

where h_{ijt} denotes this individual's health expenditures and c_{ijt} corresponds to his nonhealth consumption expenditures; Q_{jt} is the "quality of health care" per unit of health care expenditure in area j at time t. Here by "area" we refer to geographic areas approximating local health care markets. In our empirical work, we will look at economic sub-regions and then aggregate the data to state level to investigate whether some of these technology and policy responses might be more pronounced at a higher level of aggregation. The functional form in (3.1), that is, multiplicatively separable in health and nonhealth consumption, is adopted both to simplify the exposition and to link our equations to Hall and Jones (2007), whose dynamic model also has a static representation identical to (3.1). The budget constraint of the individual is written as

$$c_{ijt} + h_{ijt} \leq y_{ijt}, \tag{2.2}$$

where recall that both c_{ijt} and h_{ijt} are expenditures (and we therefore have no prices on the right-hand side; the relative price of health care is already incorporated into Q_{jt}). Assuming that both π and u are concave and differentiable, the individual's optimal demand for health expenditures leads to the following simple equation for the share of income spent on health:

$$\frac{h_{ijt}}{y_{ijt}} = \frac{\eta_{\pi_{ijt}}/\eta_{u_{ijt}}}{1+\eta_{\pi_{ijt}}/\eta_{u_{ijt}}}, \tag{2.3}$$

where $\eta_{\pi_{ijt}} \equiv Q_{jt}h_{ijt}\pi'\left(Q_{jt}h_{ijt}\right)/\pi\left(Q_{jt}h_{ijt}\right)$ and $\eta_{u_{ijt}} \equiv c_{ijt}u'\left(c_{ijt}\right)/u\left(c_{ijt}\right)$ are the elasticities of the π and u functions evaluated at the expenditure levels of individual i in area j and time t.

As emphasized by Hall and Jones (2007), we expect the share of income spent on health to increase as incomes rise if $\eta_{u_{ijt}}$ decreases more rapidly than $\eta_{\pi_{ijt}}$ with income. However, the behavior of the quality of health care in the area, Q_{jt}, also plays an important role in the evolution of health expenditures. To see this more clearly, we take logs of both sides of (2.3), then take a first-order Taylor expansion of $\log[(\eta_{\pi_{ijt}}/\eta_{u_{ijt}}) \,/\, (1+\eta_{\pi_{ijt}}/\eta_{u_{ijt}})]$ in terms of $\log h_{ijt}$, $\log Q_{jt}$ and $\log y_{ijt}$ and rearrange to write

$$\log h_{ijt} = \zeta \log Q_{jt} + \tilde{\beta} \log y_{ijt} + \xi_{ijt}, \tag{2.4}$$

where ξ_{ijt} is an error term capturing approximation errors as well as any omitted factors. In practice, we expect the error term to have a representation of the form $\xi_{ijt} = \tilde{\alpha}_j + \tilde{\gamma}_t + \tilde{\varepsilon}_{ijt}$, with $\tilde{\alpha}_j$ and $\tilde{\gamma}_t$ corresponding to systematic differences in the demand for health care across areas and over time. In equation (2.4), $\tilde{\beta}$ measures the individual income elasticity for health expenditures holding Q_{jt} constant. This is the elasticity we would measure if we could have random variation in individual incomes within an area, holding quality of health care Q_{jt} constant; it thus corresponds to what we referred to as the "partial equilibrium" income elasticity.

Since we are interested in the role that rising income has played in the rising health share of GDP, we wish to obtain estimates that incorporate the general equilibrium effects of income on health spending, which are captured in the model by Q_{jt}; in general equilibrium, income changes may affect the "quality" of health care.[10] In this context, it is important to distinguish between "local" general equilibrium effects—corresponding to the effects of changes in income in area j on health expenditures working through their effects on Q_{jt}—and national (or global) general equilibrium effects—whereby changes in national (or global) income impact health expenditures via their effect on some "frontier" quality or the aggregate of the area qualities, i.e., the aggregate of the Q_{jt}s. Examples of local general equilibrium effects of area income on area health care quality would include hospital entry and technology adoption decisions in response to changes in local income, and local health policy decisions (such as funding of public hospitals or state-level public health insurance eligibility rules) that are responsive to local area income. Examples of national or

[10]We note that our use of the term "quality" does not imply a normative assessment of the net social benefit of changes in Q_{jt}.

global general equilibrium effects of income would include the development of new technologies induced by national or global income changes and the responsiveness of national health policy decisions (such as Medicare policy) to national income.

To capture these two distinct mechanisms we write

$$\log Q_{jt} = \bar{\alpha}_j + \kappa_0 \log y_{jt} + \kappa_1 \log y_t + \lambda_1 s_t, \tag{2.5}$$

where y_{jt} is average (per capita) income in area j at time t and y_t is average national income,[11] κ_0 measures local general equilibrium effects and κ_1 captures national or global general equilibrium effects; in addition $\bar{\alpha}_j$ captures other (orthogonal to income) sources of variation in the quality of health care across areas, and s_t captures other (orthogonal to income) factors affecting the quality of health care in the aggregate, such as autonomous scientific advances. Substituting (2.5) into (2.4), averaging across all individuals within area j, and proxying the average of logs with the log of the average, we obtain

$$\log h_{jt} \simeq \alpha_j + \gamma_t + \beta \log y_{jt} + \varepsilon_{jt}, \tag{2.6}$$

where h_{jt} is average health expenditure in area j at time t, and we have $\alpha_j \simeq \tilde{\alpha}_j + \zeta\bar{\alpha}_j$, $\gamma_t \simeq \tilde{\gamma}_t + \zeta(\kappa_1 \log y_t + \lambda_1 s_t)$, and $\beta \simeq \tilde{\beta} + \zeta\kappa_0$. Note also that equation (2.6) could have been equivalently written in its "aggregate form," with the log of total area health expenditure, $\log H_{jt}$, on the left-hand side, and the log of total area income, $\log Y_{jt}$, on the right-hand side. In our empirical work, it will be more convenient to start with this version, though our main estimates will come from equations expressed in variants of per capita units as in (2.6).

Equation (2.6) emphasizes that the income elasticity β we estimate will differ from the partial equilibrium income elasticity ($\tilde{\beta}$) due to local general equilibrium effects ($\zeta\kappa_0$). For example, when the income of a single individual in an area increases, the types of health care that this individual has access to will remain constant and this may limit his willingness to spend on health care. In contrast, if the entire

[11] For induced innovation, we could also take y_t to represent average income in the OECD.

area becomes more prosperous, local hospitals may adopt new technologies or new practices that increase the willingness of (a subset of) the local population to spend on health care. Thus, while the partial equilibrium income elasticity $\tilde{\beta}$ might small, the local general equilibrium elasticity β could be substantially larger.[12]

Equation (2.6) also emphasizes that national or global general equilibrium income effects—such as induced innovation or national policy responses—are absorbed by the time effects, the γ_ts, and are thus not captured in our estimates of β. To the extent that these national or global general equilibrium income effects are quantitatively important, our estimates will understate the national or global relationship between income and health. We view this as an important but inevitable casualty of our empirical strategy, which attempts to obtain "causal" estimates of the impact of income on health expenditures. We are not aware of alternative empirical strategies that could generate convincing estimates of national and global general equilibrium effects of rising income (in particular, any pure time-series strategy would confound the effects of income with those represented by s_t in equation (2.5)). Instead, our strategy is to provide credible estimates that incorporate both partial equilibrium and local general equilibrium income effects. We then draw on supplementary evidence to try to gauge whether there are likely to be quantitatively important national general equilibrium income effects not captured by our analysis. As we discuss in Section 2.2.3 below, this supplementary evidence and analysis suggest that the national relationship between income and health expenditures is not significantly different than the relationship estimated from our empirical approach.

2.1.2 Empirical Strategy

Our empirical strategy is to estimate (2.6) using plausibly exogenous variation in income across areas. In particular, we will instrument for income in different geographic areas (approximating local economies) with time-series variation in oil prices interacted with cross-sectional variation in the oil intensity of the different local economies. We then examine the relationship between the resulting changes in in-

[12] As noted in the Introduction, β could also be smaller than $\tilde{\beta}$.

come and changes in health care spending using panel data on area-level health care spending. Let us first write (2.6) in its aggregate form and with covariates as follows

$$\log H_{jt} = \alpha_j + \gamma_t + \beta \log Y_{jt} + \mathbf{X}_{jt}^T \phi + \varepsilon_{jt}, \tag{2.7}$$

where Y_{jt} is total area income, $\mathbf{X}_{jt}$ denotes a vector of other covariates that are included in some of our specifications (and $\mathbf{X}_{jt}^T$ denotes its transpose). In our baseline specification, there are no $\mathbf{X}_{jt}$s, and H_{jt} is total hospital expenditures in area j at time t. The α_js are area fixed effects measuring any time-invariant differences across the different geographic areas. The γ_ts are year fixed effects, capturing any common (proportional) changes in health care spending each year.[13] For convenience and transparency, we begin by estimating this aggregate form of (2.6), and then turn to estimating variants of the per capita specification shown in (2.6).

The simplest strategy would be to estimate β in equation (2.7) using ordinary least squares (OLS). However, OLS estimates of β are likely to be biased. Moreover, the sign of the bias is a priori ambiguous. For example, if income is positively correlated with (unobserved) health and healthier areas have lower health care expenditures, the OLS estimates would be biased downwards. If, on the other hand, income is positively correlated with insurance coverage and insurance encourages increased health care spending, OLS estimates would be biased upwards.

Our empirical strategy attempts to isolate potentially-exogenous sources of variation in local area income, Y_{jt} (or equivalently in local per capita income, y_{jt}, in later specifications). We instrument for changes in area income by exploiting the differential impact of (global) changes in oil prices across areas of the country in which oil plays a more or less significant role in the local economy. In particular, we instrument for $\log Y_{jt}$ in equation (2.7) with the following first-stage regression:

$$\log Y_{jt} = \alpha'_j + \gamma'_t + \delta(\log p_{t-1} \times I_j) + \mathbf{X}_{jt}^T \phi' + u_{jt}, \tag{2.8}$$

[13]The specification with the dependent variable, hospital expenditures, in logs rather than in levels is attractive both because the distribution of hospital expenditures across areas is highly right skewed (see Figure 4b below) and because it implies that year fixed effects correspond to constant proportional (rather than constant level) changes in health spending across all areas.

where p_{t-1} is the global spot oil price in the previous year, and I_j is a (time-invariant) measure of the role of oil in the local economy. The α_j's and γ_t's are defined similarly to the α_js and γ_ts in equation (2.7). In our baseline specifications, I_j will be proxied by the total amount of oil reserves in area j. Throughout, we use oil prices dated $t-1$ in the regression for income at time t to allow for a lag in the translation of oil price changes into income changes. We show in the on-line Appendix (Section B) that the estimates and implied elasticities are similar when we instead use oil prices at time t. The year fixed effects in both the first and second stage will capture any common (proportional) effects of oil price changes on area income and health care expenditures that are independent of the role of oil in the local economy; these may be operating, for example, through the effects of oil prices on costs of living or production.

Our identifying assumption is that, absent oil price changes, health expenditures in areas with different oil reserves would have grown at similar rates. This is reasonable since both global oil prices and the location of oil reserves are not affected by, and should not be correlated with, changes in an area's demand for health care. Naturally, areas with different amounts of oil reserves may differ in ways that could affect health expenditures. Any such differences that are time-invariant will be captured by the area fixed effects (the α_js and α_j's) in equations (2.7) and (2.8). Only differential trends in health expenditures across these areas would be a threat to the validity of our instrumental-variables strategy. As a basic step to increase comparability across areas and to limit potential differential trends, our baseline analysis focuses on the Southern United States—which contains about 50% of the oil in the United States (Oil and Gas Journal Data Book, 2000). We show in the next subsection that areas of the Southern United States that differ in terms of the role of oil in the local economy (I_j in (2.8)) have similar levels of income and hospital expenditures at the start of our sample period (when oil prices had been relatively constant for at least 20 years). More importantly, in the on-line Appendix (Section B), we provide a variety of evidence to support our identifying assumption that there were no major differential trends in health expenditures across local economies correlated with their oil intensity.

Our baseline specification focuses on the period 1970-1990, which encompasses the major oil boom and bust, and uses economic sub-regions (ESRs) as our geographic units (local economies). We construct our ESRs by splitting the economic sub-regions produced by the Census ("Census ESRs") so that our ESRs do not straddle state boundaries. Census ESRs are commonly used geographic aggregations that were last revised for the 1970 Census; they consist of groupings of State Economic Areas (SEAs).[14] There are 247 ESRs in the United States overall, and 99 in our sample of 16 Southern states.[15] We discuss below the results of analyses at different levels of aggregation (in particular, state) and also explore the implications of expanding the analysis to include longer time periods and other parts of the United States.

2.1.3 Data and Descriptive Statistics

Estimation of equations (2.7) and (2.8) requires time-series data on oil prices, cross-sectional data on the oil intensity of the local economy, panel data on the income in each area, and panel data on health expenditures in each area. We briefly describe the construction of our main data series here. Table 1 provides summary statistics on some of our main variables.

Oil prices We measure oil prices by the average annual spot oil price from the West Texas Intermediate series.[16] Figure 2 shows the time series of average annual spot oil prices from 1950 to 2005. We focus primarily on the period 1970-1990, as these two decades encompass the major oil boom and bust. Oil prices rose dramatically over the 1970s from \$3.35 per barrel in 1970 to a high of \$37.38 per barrel in 1980. This oil boom was followed by an oil bust; oil prices declined starting in 1980 to a trough

[14]ESRs frequently cross state boundaries. In contrast, SEAs do not cross state boundaries and are defined on the basis of a combination of demographic, economic, agricultural, topographic and natural resource considerations. In metropolitan areas, SEAs are based on standard metropolitan areas (SMSAs); for SMSAs that straddle two or more states, each part becomes a separate SEA.

[15]Our baseline sample is 2065 observations instead of $99 \times 21 = 2079$ observations because of four ESR-years of missing hospital data and because Washington D.C. does not appear in the hospital data until 1980. Restricting the sample to include only ESRs that appear in all years does not affect results.

[16]These data are available at `http://research.stlouisfed.org/fred2/series/OILPRICE/downloaddata?cid=98`.

of \$15.04 per barrel in 1986. We discuss below the effects of extending the analysis to include the later oil boom that began at the end of the 1990s, and show in the on-line Appendix (Section B) the results of falsification exercises during the pre-boom 1950s and 1960s.

It is worth noting at this point that, as also documented by several other researchers (e.g., Hamilton, 2008, Kline, 2008), oil price shocks appear to be permanent. This suggests that our empirical strategy will be informative about the effects of permanent (rather than transitory) changes in income on health care expenditures. We discuss and further document this in subsection 2.3.4 below.

Oil intensity Our primary measure of the oil intensity of area j is an estimate of the total oil reserves in that area (since discovery). We draw on data from the 2000 Edition of the Oil and Gas Journal Data Book, which includes information on all 306 oil wells in the United States of more than 100 million barrels in total size. Total oil reserves are calculated as estimated remaining reserves plus total cumulative oil production as of 1998; they are thus not affected by the prior intensity of oil extraction in the area. Throughout, we refer to these as "large" oil wells. Our baseline analysis is limited to the Southern United States, which contains 161 of the 306 large oil wells in the United States and 51% of the total oil reserves of these oil wells.[17]

Figure 3 shows the cross-sectional variation in oil reserves across different areas of the South. It indicates that the importance of oil to the local economy varies substantially across different areas of the South, including substantial within state variation. For example, approximately 70 percent (69 out of 99) of the ESRs in the Southern United States have no large oil wells. Conditional on having a large oil well, the standard deviation in oil reserves across ESRs in the Southern US is more than 2500 million barrels (relative to a mean reserve conditional on having any reserves of 1700 million barrels). As a result of this variation, as we shall see, different areas

[17]According to the 2000 Edition of the Oil and Gas Data Book, there is only one large well in the South that is listed as having been discovered after 1970 (Giddings, TX in 1971). Excluding this well has no effect on our results. There are also 60 (out of the 306) oil wells that are located off-shore and thus were not assigned to any county. These off-shore wells account for 12% of the oil reserves in the data.

experienced differential changes in income in response to changing oil prices; this is the basis of our first stage.

In some of our analyses we also draw on data from the 1970 Census on the mining share of employment in 1970 to help measure oil intensity of an area. The mining share includes all workers in oil mining, natural gas and coal mining (it is not available separately for oil mining).[18]

Area income Our primary data on ESR income comes from aggregating up county-level annual payroll (for all establishments) from the County Business Patterns (CBP).[19] We also obtain ESR-level employment data from the CBP in the same manner. The CBP data are attractive for our purposes because of their level of disaggregation, enabling us to construct ESR-level measures of income. Figure 4a provides a histogram of the logarithm (log) of income from the CBP across ESRs. The distribution of log income appears to be well approximated by a normal distribution.

A potential drawback of these data is that they do not include capital income. To investigate whether the exclusion of capital income has a systematic effect on our results, we also repeat our analysis at the state level using annual data on gross state product (GSP), which includes both labor and capital income. We also use industry-specific GSP estimates as a dependent variable to provide comparative estimates of income elasticities in different industries.[20]

Area health spending Our primary data on area health spending are obtained by aggregating up hospital level data from the American Hospital Association's (AHA) annual census of all US hospitals. We use these data to construct our main dependent

[18]Mining share of employment is defined based on the 1970 Census of Population (Volume 1: Characteristics of the Population, Table 123, Parts 2-9 & 11-52).

[19]The CBP is an annual establishment survey of all establishments in the Business Register at the Census Bureau. The CBP data are available on-line at the Geospatial & Statistical Data Center at the University of Virginia for the years 1977 through 1997 (`http://fisher.lib.virginia.edu/collections/stats/cbp/county.html`) and at the U.S. Census Bureau for the years 1998 through 2006 (`http://censtats.census.gov/cbpnaic/cbpnaic.shtml`). Earlier years were hand-entered from bound volumes available at the MIT Library Storage Annex. For more information on these data see `http://www.census.gov/epcd/cbp/view/cbpmethodology.htm`.

[20]GSP data are from the Bureau of Economic Analysis (`http://www.bea.gov/regional/gsp/`).

variable, total hospital expenditures in area j and year t. Figure 4b shows a histogram of the logarithm of hospital spending from the AHA, which also has the standard shape of a normally-distributed variable.

The AHA data also contain other measures of hospital activity, which we use below to investigate which components of health expenditure respond to the rise in income and to investigate the impact of rising income on hospital technology adoption. Specifically, the AHA data contain total hospital expenditures, payroll expenditures, full time equivalent employment, admissions, inpatient days, beds, and a series of binary indicator variables for whether the hospital has a variety of different technologies. For about three quarters of the years, we also have information on the levels of full-time equivalent employment of two types of nurses in the data: Registered Nurses (RNs) and Licensed Practitioner Nurses (LPNs), which together constitute about 20% of total hospital employment. RNs are considerably more skilled than LPNs and we use the ratio of RNs to RNs and LPNs combined as a proxy for the skill mix.[21]

There are three key advantages of the AHA data. First, they are extremely high quality. Relatedly, they appear to be unique among annual sub-national data on health expenditures from our time period in that they are constructed independently each year, and therefore do not rely on some degree of interpolation between years. Second, they allow us to conduct our analysis at a level of aggregation below the state and thus to exploit the substantial within-state variation in oil intensity shown in Figure 3a. Third, they allow us to measure other components of health care activity. In particular, using these data we can measure hospital technology adoption decisions and thereby investigate potential global general equilibrium effects through induced innovation.

The major drawback of the AHA data is that they do not contain information on non-hospital components of health expenditures. To investigate whether the focus on hospital spending may lead to biased estimates of the income elasticity of total health

[21]RN certification requires about twice as many years of training as LPN certification and RNs are paid substantially higher hourly wages (see Acemoglu and Finkelstein, 2008).

expenditures, we use data from the Health Care Financing Administration (HCFA), which produces state-level estimates of total personal health care expenditures and its components, although only for a subset of our study years (Levit 1982, 1985).[22] In addition, we also examine decadal state-level Census data on the earnings of various groups of health care providers.

Population To investigate the extent of migration in response to our income variation, we use annual data on total area population and on area population by five year age groups from the Current Population Reports (CPR). Crucially, for our purposes, population is not interpolated between censuses but rather is imputed annually based on a variety of administrative data sources including data on births, deaths, school enrollment, and tax returns (US Census Bureau, various states and years, and Siegal, 2002).[23]

Finally, to gauge the relative intensity of hospital use among individuals of different age groups, we use data on the age profile of hospital use constructed from the National Health Interview Survey (NHIS), which we pool between 1973 and 1991.

Comparison across areas with different oil intensity Table 2 examines whether there are significant differences in income and various measures of hospital activity in 1970 across ESRs with different levels of oil reserves. We look at this relationship in our baseline sample of the 16 Southern United States. Columns 3 and 4 of this table show that there is no statistically or economically significant relationship between oil reserves and any (or all) of population, total employment, hospital ex-

[22] As we discuss in more detail in Section 2.3 below, the HCFA data are constructed partly based on interpolation between years, which is an important caveat for regression analysis based on these data. Data from 1972 and 1976-1978 were obtained from Levit (1982, 1985). Data for 1980-1990 were obtained from the Centers of Medicare & Medicaid Services on-line at `http://www.cms.hhs.gov/NationalHealthExpendData/05_NationalHealthAccountsStateHealthAccountsResidence.asp`. The data include total health expenditures and expenditures on the following components (which sum to the total): Hospital Care, Physicians' Services, Dentists' Services, Drugs and Other Medical Nondurables, Eyeglasses and Appliances, Nursing Home Care, and Other Health Services (which include Home Health Care, Other Professional Services, and Other Personal Services).

[23] The Current Population Reports data are available on-line at the U.S. Census Bureau (`http://www.census.gov/popest/archives/pre-1980/` and `http://www.census.gov/popest/archives/1980s/`).

penditures, hospital beds and total income. In each case, the association with oil reserves is statistically indistinguishable from zero and the magnitude of variation is small (one standard deviation change in oil reserve is associated with only about one tenth of one standard deviation change in each of these variables). This offers some preliminary support for our exclusion restriction that, absent the oil price changes in the 1970s and 1980s, ESRs with different levels of oil reserves would have been on similar trends in terms of their hospital expenditures and utilization. The on-line Appendix (Section B) provides a much more systematic investigation of the validity of our exclusion restriction.

2.2 Main Results

2.2.1 First Stage

Table 3 shows the relationship between ESR income and our instrument. The first column shows the results from estimating equation (2.8). In this and all subsequent estimates, we allow for an arbitrary variance covariance matrix within each state.[24] The results in column 1 indicate a positive and strong first stage: ESRs with greater oil reserves experience greater changes in income in response to oil price changes than areas with less oil. The F-statistic is 18.74. We defer a discussion of the magnitude of the first stage until a little later in this section.

To examine the sources of the increase in income, column 2 re-estimates the first-stage equation (2.8) using log area employment on the left-hand side instead of log area income. The results indicate that areas with more oil also experience greater change in employment when oil prices change. The coefficient on our instrument, δ, is of approximately the same magnitude in columns 1 and 2, suggesting that all (or most) of the changes in income associated with oil price movements across areas with different levels of oil reserves may be due to changes in employment at constant

[24]Because of concerns of the small sample properties of clustering with only 16 states, we experimented with alternative small sample corrections, as well as alternative strategies to correct for potential serial correlation. The alternative procedures produce similar results, and are discussed in the on-line Appendix (Section B).

wages. This is consistent with our prior expectations that oil workers should be close substitutes to other workers and have a relatively elastic labor supply in the local labor market. It is also consistent with the stylized fact that labor income changes at short-run frequencies (e.g., over the business cycle) are largely driven by employment changes, with little movements in wage per worker.[25] In contrast to our source of income variation, about half of the growth in income between 1960 to 2005 is due to increased employment, while the other half is due to increased wages per employee (US Census Bureau, 2008). In Section 2.3, we discuss the possible implications of extrapolating from our income changes to the effects of the secular increase in incomes in the US economy.

The impact of our instrument on employment and existing evidence on migration responses to local economic conditions (e.g., Blanchard and Katz, 1992) suggest that our instrument may also affect area population. Any increase in population in high oil areas relative to low oil areas may increase health expenditures directly, potentially over-stating the effect of increased income on hospital spending among a (constant) population. Column 3 explores this issue by re-estimating equation (2.8) with log population as the new dependent variable. The results indicate that our instrument also predicts population, so that part of the increase in area income we estimate reflects increases in area population; a comparison of columns 2 and 3 suggests that about one third of the effect of the instrument on employment can be accounted for by its effects on population.

A natural solution is to convert both income (our endogenous right-hand side variable) and hospital expenditures (our dependent variable of interest) into per capita terms, so that the structural equation focuses on the impact of income per capita on hospital spending per capita (the same instrument now used for income per capita in the first stage); this also matches more closely our estimating equation (2.6) from the motivating theory. The first-stage results from estimating equation (2.8) with log income per capita on the left-hand side are shown in column 4. Consistent with

[25]See, for example, Abraham and Haltiwanger (1995). This does not imply that the wage per efficiency unit of labor is constant, since there may be composition effects (see, Solon, Barsky and Parker, 1994).

a comparison of columns 1 and 3, the per-capita specification shows a statistically significant but smaller first-stage effect than unadjusted specification in column 1. In particular, the first-stage coefficient is smaller than that in column 1 by 5 log points or by about 40 percent.

While the per capita specification is natural, it may in turn understate the effect of increased income on hospital spending because the population changes associated with our instrument are from disproportionately low users of hospital care. This can be seen in columns 5 and 6, in which we estimate equation (2.8) using as the dependent variable the log of the total population under 55 and the log of the total population 55 and over, respectively. The results indicate that the population response to our instrument is concentrated among the non-elderly (those under 55). In fact, it appears that the population response is concentrated among those younger than 45 (not shown in Table 3 to save space). Younger individuals consume disproportionately lower amounts of hospital care than the elderly. To illustrate this, Figure 5 shows the average annual number of hospital days for individuals in five-year age brackets estimated from the National Health Interview Survey (NHIS), pooled between 1973 and 1991. The under 55 average 0.6 hospital days per year, while individuals aged 55 and older average 2.3 hospital days per year. As a result, even though the 55 and older are only 23% of the population, they consume 38% of hospital days.

To obtain more accurate estimates of the impact of rising incomes on health expenditures (and, if anything, to err on the side of over-estimating, rather than under-estimating, income elasticities), in our baseline analysis we correct for the changes in the composition of the population rather than simply using per capita estimates. In particular, we construct a measure of "hospital utilization weighted population" in area j in year t, denoted by $HUWP_{jt}$. This measure is computed as the inner product of the vector of populations in each five year age bin in area j and year t (pop_{ajt}) with our estimate of the national average of hospital days used by that age bin ($hospdays_a$) from the pooled 1973-1991 NHIS. Namely:

$$HUWP_{jt} = \sum_a pop_{ajt} \times hospdays_a \tag{2.9}$$

Our preferred specification adjusts (i.e., divides) income in both the structural equation (2.7) and the first-stage equation (2.8) and hospital expenditures in the structural equation (2.7) by $HUWP_{jt}$ as constructed in equation (2.9). This leads to our baseline structural equation, closely resembling our motivating theoretical equation, (2.6):

$$\log \tilde{h}_{jt} = \alpha_j + \gamma_t + \beta \log \tilde{y}_{jtjt} + \mathbf{X}_{jt}^T \phi + \varepsilon_{jt}, \tag{2.10}$$

and our baseline first-stage equation:

$$\log \tilde{y}_{jt} = \alpha'_j + \gamma'_t + \delta'(\log p_{t-1} \times I_j) + \mathbf{X}_{jt}^T \phi' + u_{jt}, \tag{2.11}$$

where adjusted income ($\tilde{y}_{jt}$) and adjusted hospital expenditure ($\tilde{h}_{jt}$) are defined as

$$\tilde{y}_{jt} \equiv \frac{Y_{jt}}{HUWP_{jt}} \text{ and } \tilde{h}_{jt} \equiv \frac{H_{jt}}{HUWP_{jt}}.$$

Intuitively, both income and hospital expenditures (or other outcomes) are adjusted for hospital-use weighted population (HUWP) to capture any direct effect of our instrument on hospital-use weighted population.

The estimates of the first-stage coefficient, δ', from equation (2.11) are shown in column 7. Its magnitude lies (mechanically) in between the first-stage estimates without any migration adjustment (column 1) and with the per capita adjustment (column 4). In practice, the magnitude is about one third of the way from the per capita adjustment to the unadjusted specification. The IV estimate of the effect of income on hospital spending using the hospital utilization weighted population adjustment should therefore similarly lie in between the unadjusted estimates and the per capita adjusted estimates (and we find below that it does).

In what follows, we take the estimates from equations (2.11) and (2.10), which correct for the age-adjusted hospital utilization of the population, as our baseline/preferred specification. Because even conditional on age migrants may be healthier than the general population, the estimate of β from (2.10) might understate the effects of income on health expenditures. We therefore also report results without any adjustment for migration as well as results using the per capita adjustment. One might consider the unadjusted estimates as an upper bound on the income elasticity, and the per capita adjusted estimates are a lower bound (provided that the marginal migrant into a high-oil area in response to an oil price increase is "healthier" than the average population in the area, which seems like a reasonable assumption).[26] In practice, we will see below that these "bounds" on the income elasticity are relatively tight.

Finally, column 8 shows the HUWP-adjusted first stage but now aggregated to the state level (rather than the ESR level as in column 7); the first stage is robust to aggregation to the state level (F-statistic = 24.05).[27]

To gauge the magnitude of the first stage, we calculated that in our preferred specification (column 7) the oil price change from 1970 to 1980 is associated with a 3.6 percent larger increase in area income in areas with a one standard deviation larger amount of oil. The first stage in our preferred specification has an F-statistic of 16.58.

2.2.2 Income Elasticity of Hospital Spending and Components

Table 4 presents our central estimates of the impact of income on hospital expenditures. Column 1 reports the OLS estimate of equation (2.10) in which both hospital expenditures and income are adjusted for HUWP. The estimate of β in (2.10) is -0.027 (standard error = 0.074). This indicates that when income in an area increase by 10 percent, hospital expenditures fall by about 0.3 percent. This relationship is statistically indistinguishable from zero. As previously discussed, the OLS correlation

[26]This last presumption is both intuitive and consistent with the fact that migration is concentrated among younger individuals (see Table 3).

[27]Although the first stage is robust to aggregating up from ESR to state, it is not robust to disaggregating the data to a lower level of aggregation than the ESR (not shown). For example, we explored analyses conducted at the level of the State Economic Area (SEA); there are 194 SEAs in our sample of Southern States compared to 99 ESRs. The major concern with the SEAs is that some of them are closely linked to each other economically and residentially, thus would not be experiencing independent income variation. In this case, we would expect a significant amount of attenuation in the first stage. Consistent with this expectation, the first stage becomes weaker, with an F-statistic of only 2.06 at the SEA level. As a result, we do not report IV estimates for lower levels of aggregation.

between income and hospital spending may be biased in either direction relative to the causal effect of income on hospital spending. Our subsequent analysis suggests that in our setting the OLS estimate is downward biased.

Column 2 shows the results from the reduced form corresponding to (2.10) and (2.11) (without covariates):

$$\log \tilde{h}_{jt} = \alpha''_j + \gamma''_t + \delta''(\log p_{t-1} \times I_j) + \varepsilon''_{jt}. \tag{2.12}$$

This reduced-form estimation shows a positive and statistically significant relationship between our instrument and log hospital expenditures.

Column 3 presents our baseline IV estimate of equation (2.10). The estimated elasticity of health expenditure with respect to income is 0.723, with a standard error of 0.214.[28]

Columns 4 and 5 show IV results without any population adjustment and with a per capita population adjustment, respectively, to both hospital expenditures and income. As discussed in Section 2.2.1, these estimates can be interpreted as upper and lower bounds on the income elasticity of hospital spending. In both alternative specifications the income elasticity ranges between 0.665 and 0.801, suggesting that these bounds are reasonably tight.

The last column of Table 4 reports the results from our baseline, HUWP-adjusted specification (from column 3) but now aggregated to the state level. We estimate an income elasticity at the state level of 0.550 (standard error = 0.230). The point estimate at the state level is similar to our estimate at the ESR level of 0.723 (see column 3). We provide a more detailed discussion of state-level results in Section 2.3 but note here that, among other things, the state-level estimates allow us to capture potential general equilibrium effects, such as political economy effects, that may be more likely to occur at the level of the state than at the sub-state ESR.

Table 5 investigates which components of hospital expenditures are affected by

[28]Since we have only one instrument and one endogenous right-hand side variable, the point estimate in the IV specification can also be obtained by dividing the reduced-form estimate in column 2 by the first-stage estimate from column 7 of Table 3.

income changes. It reports the results from IV estimation of equation (2.10) using different hospital outcomes as the dependent variable.[29] Several interesting findings emerge. First, the results in columns 1 and 2 suggest that the impact of income on hospital payroll expenditures (which are about one half of total hospital expenditures) can explain all of the effect of income on total hospital expenditures. There is no evidence in column 3 of an economically or statistically significant effect of income on hospital employment. This suggests that the increase in payroll expenditures comes from a combination of an improvement in the quality of employees and/or a bidding up of the wages of (quality-adjusted) employees.

Second, we find evidence of economically and statistically significant skill upgrading associated with increased income. Column 4 shows an increase in the skill composition of employment, proxied by the ratio of skilled nurses (RNs) to all RNs and LPNs.[30] This does not rule out wage (price) effects, but suggests that at least some of the increase in payroll expenditures in column 2 comes from quality improvements. More importantly, evidence of skill upgrading also suggests that our empirical strategy is able to uncover (at least some) local general equilibrium effects; skill upgrading of hospitals is likely to be a response to the ESR-level increase in the demand for hospital services.

Third, we find no evidence that rising income is associated with an increase in hospital utilization (as measured by either admissions or patient days) or in hospital capacity (as measured by beds). These results are shown in columns 5 through 7.

[29]As detailed in the notes to Table 5, we adjust both the dependent variable and income for hospital-utilization weighted population (HUWP) to account for population migration in response to our instrument. The exceptions are in columns 4 and in columns 8-11 in which income is still adjusted for (i.e., divided by) HUWP, so that we are measuring the increase in income per adjusted population, but the dependent variable is not adjusted for HUWP. In column 4 the dependent variable is a ratio (of skilled nurses to total nurses) which would not increase mechanically with population; in columns 8-11, the dependent variables (number of hospitals, number of technologies, or indicator for specific technologies) are count variables or indicators, which would not be expected to scale linearly with population in the same way as, e.g., spending or admissions are likely to. For these reasons, we do not adjust these dependent variables for population. As discussed above, not adjusting for migration could be interpreted as providing upper bound estimates of responsiveness to income.

[30]We only have information on RN and LPN employment for the following years: 1970, 1972, 1974, 1976, 1978, 1980-2005. Our baseline elasticity estimate for hospital expenditures declines to 0.449 (s.e. 0.181) when the odd years in the 1970s are excluded.

The point estimates are uniformly negative. For admissions and patient days, the estimates are statistically significant, though this is far from a robust result. In the robustness analysis in the on-line Appendix (Section B), we document that, in contrast to the other statistically significant results in Table 5 which are highly robust, both the statistical significance of the estimated declines of admissions and patients days and the sign of their point estimates vary across specifications; in addition, the coefficient on beds changes sign (and is rarely statistically significant) in alternative specifications.[31] We therefore interpret these results as showing no response of hospital utilization or capacity to changes in income. This pattern is consistent with the time-series evidence suggesting that hospital utilization has not been increasing as incomes have risen; indeed, age-adjusted admissions rates appear to been roughly constant since 1960, while length of stay has fallen (Newhouse, 1992).

The remaining columns of Table 5 document the impact of rising income on hospital entry and technology adoption; we discuss these results in the next subsection.[32]

2.2.3 National and Global General Equilibrium Effects

As noted at the outset, our empirical strategy is designed to capture (and, as indicated by the skill upgrading results in Table 5, *does* capture) general equilibrium effects that occur at the level of the local economy. However, a thorough empirical examination of the role that rising income plays in the growth of the health care sector requires incorporating any general equilibrium effects of income on health care spending that occur at the national or global level. Two such effects that could potentially increase

[31] See in particular Appendix Table A6 for a summary of the results of the robustness analysis for these variables.

[32] We also explored the relationship between our income variation and public funding of health care, using data from the Regional Economic Information System; these data are available at the ESR level annually for our entire study period. Public spending on health care appears to fall as income rises, with Medicaid spending falling substantially more than Medicare spending. Since the income of either Medicare or Medicaid beneficiaries should not be affected much by our instrument (the former are predominantly retirees with a pre-determined income stream and the latter are, by definition, constrained to be very low income), these results likely reflect a potential crowding out of scarce hospital resources from those whose incomes have risen and perhaps also policy responses of state governments to changing incomes. The decline in Medicaid spending may further reflect reductions in eligibility for Medicaid resulting from the increase in employment. These results are available upon request.

the income elasticity of health expenditures above what we have estimated are induced innovations (which could occur at the national or global level) and national political economy responses to rising income. In this section, we explore each of these potential mechanisms in turn.

Endogenous technology responses While our estimates incorporate the impact of income on technology adoption and entry of new hospitals at the ESR level, they may understate the effects of rising incomes if these induced the development of major new global technologies, which then led to a sizable expansion in health expenditures. This concern is particularly important since technological change in health care is commonly believed to be one of the key drivers of rising health care expenditures (e.g., Newhouse, 1992, Fuchs, 1996, Congressional Budget Office, 2008).

In this subsection, we argue that an induced technology response to rising income is unlikely to have contributed to the increase in the health share of GDP. Our argument has two parts. First, if present and economically significant, an induced innovation response to rising income should also manifest itself at the ESR level in the form of entry of new hospitals (which presumably embody new technologies) and/or adoption of new technologies at existing hospitals. In particular, even though innovations take place at the national or global level, the same mechanism leading to induced innovations at the national or global level should also lead to faster adoption of these technologies in areas with greater increases in demand (e.g., Acemoglu, 2002, 2007). Second, existing theory suggests that induced innovations should be directed to sectors that are otherwise expanding rapidly (see in particular our online Appendix, Section A), while our estimates suggest that, all else equal, health expenditures increase less than proportionately with income.

Turning to the first component of the argument, we find no evidence that rising income is associated with an increase in hospital entry or technology adoption. These results are summarized in columns 8 through 11 of Table 5. Column 8 of this table shows a negative and statistically insignificant impact of income on the number of hospitals (so that the number of hospitals appears to have grown relatively more in areas experiencing slower income growth).

The rest of Table 5 turns to technology adoption. The AHA data contain binary indicators for whether the hospital has various "facilities", such as a blood bank, open heart surgery facilities, CT scanner, occupational therapy services, dental services, and genetic counseling services. These data have been previously used to study technology adoption decisions in hospitals, and in particular hospital responsiveness to economic incentives including the insurance regime and relative factor prices (see, e.g., Cutler and Sheiner, 1998, Baker and Phibbs, 2002, Finkelstein, 2007, Acemoglu and Finkelstein, 2008). Since they contain only indicator variables for the presence of various facilities, we cannot investigate the potential upgrading of existing technology or the intensity of technology use, but we can study the impact of changes in income on the total number of facilities, proxying for technology adoption decisions on the extensive margin.

During the time period we study, the AHA collects information on the presence of 172 different "facilities". These are listed, together with their sample means (the fraction of ESRs each technology is in) and the years in which they are available in Appendix Table A1. On average, a given facility is reported in the data for 7 out of the possible 21 years; only nine of the technologies are in the data for all years. Moreover, as is readily apparent from Appendix Table A1, the list encompasses a range of very different types of facilities. Given these two features of the data, we pursue two complementary approaches to analyzing the relationship between income and technology adoption with the AHA data.

Our first approach to investigating the impact of income on technology adoption, which is shown in column 9, treats all facilities equally and measures technology as the log of the number of distinct technologies in a given ESR in a given year. The year fixed effects in our IV estimate of equation (2.10) adjust for the fact that the set of technologies reported in each year differs. The results show no substantively or statistically significant evidence of an increase in the number of distinct technologies in the area in response to the increase in income. The point estimate on income is negative and statistically insignificant. It is also substantively small, suggesting that a 10 percent increase in area income is associated with a statistically insignificant decrease in the number of technologies in the area of 1.3 percent.[33]

A drawback of this approach is that it treats all technologies as perfect substitutes. As an alternative, we estimated hazard models of the time to adoption for specific technologies that are in the data for at least 15 years of our 21 year sample period. As in Acemoglu and Finkelstein (2008), we limit our analysis to technologies that were identified as "high tech" by previous researchers (Cutler and Sheiner, 1998, Baker, 2001, and Baker and Phibbs, 2002). Unfortunately, there are only two technologies that meet these criteria in our sample: open heart surgery and diagnostic radioisotope facility. Both have been found in other work to be responsive to economic incentives (Finkelstein, 2007, Acemoglu and Finkelstein, 2008). Both of these technologies were diffusing over our sample period, though open heart surgery started from a lower prevalence and diffused more rapidly.[34] To investigate the impact of ESR income on local technology adoption decisions, we estimate semi-parametric Cox hazard models for these two technologies as functions of income. In particular, the conditional probability that ESR j adopts the technology in question at time t (meaning that at least one hospital in the ESR adopts the technology conditional on there being no hospital in the area that had previously adopted this technology) is modeled as

$$\lambda_{jt} = \lambda_{0t} \exp(\beta \log \tilde{y}_{jt} + \mathbf{X}_j^T \phi), \tag{2.13}$$

where λ_{0t} is a fully flexible, non-parametric baseline hazard, $\tilde{y}_{jt}$ is our baseline measure of (HUWP-adjusted) income, and $\mathbf{X}_j$ is a vector of (time-invariant) covariates. Since we have at most a single transition (adoption) for each ESR, we cannot include ESR fixed effects in the hazard model. Instead, we include time-invariant ESR

[33]To provide some context for comparison, using the same technology measure (but at the hospital level rather than at the ESR level) Acemoglu and Finkelstein (2008) show that, in its first three years, the introduction of Medicare PPS was associated with, on average, the adoption of one new technology at the hospital level (about a 4 percent increase in the average number of distinct technologies that the hospital has).

[34]Open heart surgery is in our data for all 21 years (1970-1990) and diagnostic radioisotope therapy for 19 years (1972-1990). Only 43 percent of ESRs had open heart surgery technology in 1970, whereas about three quarters of ESRs did so by 1990. About three quarters of ESRs had diagnostic radioisotope faciltiies in 1972 and 92 percent had it by 1990.

characteristics in the vector $\mathbf{X}_j$, in particular, region fixed effects for the three census regions within the South, total hospital expenditures in 1970, and total hospital beds in 1970. The fully flexible baseline hazard in the Cox model is specified with respect to calendar time and thus controls for time effects. As in our baseline specification, income is an endogenous right-hand side variable, which we instrument with $\log p_{t-1} \times I_j$. We implement our instrumental variables estimator using a control function approach (Newey, Powell, and Vella, 1999). Specifically, we include the residual ($\hat{u}_{jt}$) from the first-stage regression in equation (2.11) as an additional covariate in equation (2.13). We report bootstrapped standard errors and p-values for this two-step estimator. The results reported in columns 10 and 11 in Table 5 show no evidence of a significant increase in technology adoption associated with an increase in income. The point estimates suggest a negative relationship between log income and adoption of open-heart surgery, and a positive relationship between log income and adoption of the diagnostic radioisotope facility. However, both estimates are imprecise and not statistically different from zero.[35]

Next, turning to the theoretical argument, the on-line Appendix (Section A) outlines a simple model of induced innovations and demonstrates that development of new technologies will tend to be directed toward sectors that are expanding more rapidly. The implications of this theory are consistent with existing empirical evidence, which indicate that medical innovation responds to expected market size (e.g., Acemoglu and Linn, 2004, Finkelstein, 2004). In the present context, these theoretical expectations imply that innovations induced by the secular rise in incomes should not be favoring the health care sector. In particular, our point estimates suggest that, ignoring induced technology effects, health care expenditures increase less than proportionately with aggregate income. Thus, as incomes rise, the market size for health care technologies will increase less than the market size for a range of other technologies. As a consequence, the induced technology channel suggests that there

[35]By contrast, Acemoglu and Finkelstein (2008) find statistically significant increases in the adoption of both of these technologies in response to a change in Medicare's hospital reimbursement policy for labor inputs. This suggests that the adoption of these technologies is generally responsive to economic incentives.

should not be disproportionate technological advances in the health care sector in response to the secular increase in incomes. As the model in the on-line Appendix highlights, the main exception to this conclusion is that even a less than proportionate increase in the size of the market for health care technologies might jump-start medical technological advances if technological change in the health care sector was unprofitable prior to income reaching a certain minimum threshold. This exception seems implausible (at least to us) given that advances in medical technologies have been ongoing for more than a century and plausibly at roughly a constant rate (as mortality has been declining at a roughly constant rate over this same period, e.g., Cutler and Meara, 2003).[36]

Limited income-induced technology effects for the health care sector are also consistent with the results reported in Table 5, which show no significant effects on hospital entry or technology adoption driven by ESR-level income changes. The lack of a response in hospital entry and technology adoption bolsters the argument that, because the relative market size for the health care sector does not increase disproportionately following an increase in income, the induced technology effects should also be limited.

Overall, while we cannot conclusively rule out major national or global induced technology responses to the secular increase in income in the United States, which could in turn have further effects on health expenditures, our empirical evidence and theoretical expectations suggest that these effects should be relatively small and thus should not change our basic conclusion that rising incomes are unlikely to be the major factor in the run-up in the share of GDP spent on health care.

Political economy effects of rising incomes Although our empirical strategy would not capture any effect of income on health care expenditures that operate via a national political economy response to rising income, our state-level results incor-

[36]Of course, the specific nature of medical technological progress has varied over time. For example, improvements in sanitation and other public health measures were a primary factor in mortality declines early in the 20th century, while penicillin and other antibiotics were a key factor mid-century, and medical interventions that reduce cardiovascular disease mortality were critical in the latter part of the century (Cutler and Meara, 2003).

porate potential responses at the state and sub-state levels. The similarity between the estimate of the income elasticity at the state level and at the ESR level (compare columns 6 and 3 of Table 4) suggests that these state-level policy responses do not significantly increase the responsiveness of health care expenditures to income, although there may be substantial sub-state level policy responses captured by both our ESR- and state-level estimates.[37]

While our empirical strategy does not incorporate national political economy effects resulting from rising incomes, health policy in the United States is highly decentralized, with much of the public involvement occurring at the state (or lower) level of government. Therefore our empirical strategy likely captures much of the potential political economy responses. This holds for both public provision and public financing of health care, both of which are potentially affected by changes in income.

In terms of public provision of health care, about one third of hospitals in the United States (accounting for about one third of hospital expenditures) are publicly owned. About 85 percent of these hospitals (constituting about three-quarters of public hospital expenditures) are non-federal (i.e., state-, county-, or city-owned). Thus most of any effect that income changes have on public support for hospital financing would be incorporated into our state-level analysis.

In terms of public financing of health care, by far the two largest sources are Medicare and Medicaid, which have similar levels of spending (CMS, 2006). Medicaid is jointly financed by the federal and state governments but the states are given considerable autonomy in the design of program eligibility and benefit requirements (Gruber, 2003). Political economy effects of changing income on Medicaid design are likely to be captured by our estimates using state-level variation.

Medicare, in contrast, is a fully federal program, so that any political economy effects of income on Medicare design would not be captured by our estimates. This is a potentially important channel through which rising income may affect health

[37]This observation also underscores that, as already emphasized in the Introduction, the empirical relationship between income and health spending in the United States in the latter half of the 20th century, which we are exploring, may reflect a variety of institutional factors beyond the willingness of households to spend more on health care as their incomes grow.

spending, and not one that we can directly estimate. Nevertheless, it is reassuring in this regard that Medicare spending per beneficiary over our time period has not risen faster than overall health spending per capita.[38] If rising income had quantitatively important national political economy effects in terms of Medicare generosity, one might expect to see Medicare spending per beneficiary growing faster than overall health spending per capita as incomes have risen. The fact that it has not offers some suggestive evidence that any potential political economy responses to rising incomes working through Medicare does not introduce a serious downward bias in our estimate of the role of income growth in the run-up of the health share of GDP.

2.3 The Role of Income in Rising Health Share of GDP

We now present the implications of our estimates for the role of rising income in explaining the rising health share in the United States. The bulk of the section is then devoted to a discussion of several potential concerns and caveats with this out-of-sample extrapolation exercise.

2.3.1 Income and the Rising Health Share of GDP

Let us focus on the results from our baseline specification (Table 4, column 3), which are roughly in the middle of the range of elasticities we report in various alternative specifications below.

The point estimate of an elasticity of 0.72 implies that the approximate doubling of real per capita GDP between 1960 and 2005 (from \$19,212 to \$41,874 in \$2005) should have caused a *decline* in the health share of GDP from 5 percent to about

[38]We compared the growth in per capita health expenditures to the growth in per beneficiary Medicare spending from 1975 to 2005. We started in 1975 to allow the Medicare program (which only began in 1965 and expanded to cover SSDI recipients starting in 1973) to be fully phased in. Between 1975 and 2005 Medicare spending per beneficiary grew at an average annualized rate of 7.86%, while health spending per capita grew at 7.62%. Data on total and Medicare health expenditures and Medicare beneficiaries can be found at `http://www.cms.hhs.gov/nationalhealthexpenddata/` and `http://www.cms.hhs.gov/MedicareEnRpts/`.

4 percent. The upper end of the 95 percent confidence interval from our baseline estimate is an income elasticity of 1.13. This allows us to reject a role of rising income in increasing the health share of GDP by more than 0.5 percentage points between 1960 and 2005, i.e., it does not explain more than 5 percent of the overall increase in health share over this time period.

We can also interpret our estimates in terms of their implications for rising income in explaining rising health expenditures (rather than the rising health share of GDP). The point estimate suggests that rising real per capita income may be able to explain about 15 percent of the rise in real per capita health expenditures, while the upper end of the 95 percent confidence interval allows us to reject a role for rising real per capita income in explaining more than one quarter of the rise in real per capita health spending.[39]

Therefore, our results suggest that while rising income may be an important component of growing health expenditures, it is unlikely to have contributed much to the increase in the *share* of GDP spent on health care in the United States. We next turn to several potential concerns with this extrapolation exercise.

2.3.2 Hospital Spending Versus Total Health Expenditure

An important limitation of our estimates is that the dependent variable measures hospital expenditures rather than total health expenditures, which may have different income elasticities. Hospital expenditures are the single largest component of health care expenditures, accounting for close to two-fifths of the total. By contrast, spending on physicians accounts for about one fifth of total health expenditures, and spending on drugs accounts for about one-tenth; these shares have been roughly constant since 1960 (CMS, 2006).

Our reading of the available evidence is that total health expenditures are unlikely to have a significantly higher income elasticity than hospital spending. The first piece

[39] On the basis of the existing correlation studies (described in the Introduction), past studies that have attempted to decompose the causes of the rise in health spending have concluded that the rise in income may account for anywhere from 5 percent (Cutler, 1995) to a quarter (Newhouse, 1992) of the spending growth.

of suggestive evidence comes from Figure 1, which shows that the hospital share of total health expenditures has been roughly constant over the last half century. If income elasticities were higher for the non-hospital components of health expenditures, and if the rise in income over this time period were the major driver of the increase in health expenditures, we should see (all else equal) a decline in the share of hospital spending in overall health expenditure. The fact that Figure 1 shows no such decline supports our overall conclusion.

Our second piece of evidence comes from estimates of income elasticities of overall health care expenditures and of the hospital- and non-hospital components thereof, based on several complementary data sources. We use these data to investigate whether there is any evidence that overall health expenditures are more responsive than hospital expenditures to changes in income. To preview, although estimates from the other available data sources are often quite imprecise (motivating our preference for the AHA data set), we do not find any evidence that overall health expenditures are more income elastic than hospital expenditures.

We have state-level data on total health expenditures and its components from the Health Care Financing Administration (HCFA) for 1972, 1976-1978 and 1980-1990 (instead of our baseline sample 1970-1990). The HCFA estimates are based on a combination of administrative and survey data. An important problem with these data is that each component is interpolated whenever data are missing between years (Levit, 1982, 1985). Such interpolation may bias the estimated coefficients, so the results from this data set have to be interpreted with caution.

Table 6 presents estimates from the HCFA data. Since we lose some variation by aggregating from the ESR level to the state level, we report results both for our baseline sample of the 16 the Southern states (Panel A) and for the entire United States (Panel B). Column 1 shows that our first stage is robust to state-level analysis for the subset of years for which we have HCFA data. Columns 2 and 3 show our estimated income elasticity from the HCFA data for total health expenditures and the hospital subcomponent, respectively. Both estimated income elasticities are positive but quantitatively small and imprecise, and thus statistically insignificant.

The income elasticity of hospital spending using the HCFA data is also noticeably smaller than that estimated using the AHA data.[40]

However, most importantly for our purposes, the point estimates in columns 2 and 3 of Table 6 suggest similar income elasticities for hospital expenditures and total health expenditures. Columns 4 through 9 present results for the other components of health expenditures, and provide some intuition for why hospital and total health expenditure income elasticities may be similar. The point estimates suggest that the income elasticities of spending on physician services, on dental services, on drugs and other medical non-durables, and on vision products are greater than the income elasticity of hospital spending, while nursing home care and other health services have large negative income elasticities.[41] Overall, the results in Table 6 are generally imprecisely estimated, but the point estimates are uniformly consistent with similar income elasticities for total health expenditures and for hospital expenditures.

Results from several other data sources are also consistent with this conclusion, though again are similarly imprecise. We examined the income elasticity of state-level Health Services Gross State Product (GSP) from 1970-1990. Health services GSP account for roughly 26% of total health expenditures. Our estimates using health services GSP show no evidence of a greater income elasticity than that for hospital spending; indeed the point estimates are considerably smaller than our estimates for hospital expenditures, although they are quite imprecise.[42]

[40]The hospital expenditure data in the HCFA series are estimated using the AHA data for non-federal hospitals, but use unpublished Federal agency data for federal hospital expenditures (Levit, 1982). There are also several differences between how we use the AHA data and how they are used in creating the HCFA data. Most importantly, the HCFA estimates interpolate missing data (Levit, 1982, 1985). Average state-year hospital expenditures are similar in the two data sets ($2,641 million from the HCFA data compared to $2,333 million for the same state-years in the AHA data). Log hospital expenditures are also highly correlated across the two data sets at the state-year level (correlation = 0.98). However, conditional on state and year fixed effects, the correlation in the residual log hospital expenditures is only 0.67. This presumably helps explain why the income elasticity estimates differ. Using our AHA hospital data at the state level for the full United States and limiting the sample to the years for which the HCFA data are available (i.e., the analog of Table 6 column 3 panel B), we estimate a statistically significant income elasticity of 0.509 (standard error = 0.225). This is statistically indistinguishable from the HCFA estimate of 0.139 (standard error = 0.151).

[41]The large negative income elasticity for nursing home care strikes us as intuitive. Wealthier individuals can more easily pay for assistance at home to substitue for nursing home care (which Medicaid will cover) than can poor individuals.

[42]The results for state-level Health Services GSP are shown in Table 8, column 6, Panels A and

We also examined the impact of area income on the income of different groups of health care providers (results available on request). If non-hospital components of health care expenditures—such as physician expenditures—are substantially more income elastic than hospital expenditures, we would expect to find that the earnings of the non-hospital based health care providers are also substantially more income elastic than hospital expenditures and than the earnings of health care providers that contribute to hospital expenditures, such as nurses and health care technicians. Using decadal Census data aggregated to the state level, we estimated the income elasticity of the earnings of the following groups of health care providers: physicians, nurses, health care technicians (including clinical laboratory technicians and therapy assistants), and other health services workers (including health aids, nursing aids and attendants).[43] Our IV point estimates show no evidence that physician earnings are more responses to area income than hospital expenditures or than the earnings of other health care providers. However, the estimates using the Census income data—particularly those for physician income—are noticeably less precise than those from comparable specifications using the AHA data on hospital expenditures, so that one should not place too much emphasis on these results.[44]

Overall, while there are important limitations to each data source, a number of complementary data sets with information on state-level health expenditures suggest

B). The rest of that table is discussed in subsection 2.3.4 below. Comparable state-level estimates for hospital expenditures are show in Table 7, Panel A, columns 1 and 3. The Gross State Product (State GDP) estimates are produced annually by the Bureau of Economic Analysis. The specific industries within health services (SIC code 80) are listed at `http://www.census.gov/epcd/naics/NSIC8B.HTM#S80`. The major source of state data for the health services GSP estimates are sales and payrolls from the (quinquennial) census of service industries; intercensal years are interpolated and extrapolated using wages and salaries reported annually to the BEA (see `http://www.bea.gov/regional/pdf/gsp/GDPState.pdf`).

[43]Our first stage is robust to aggregation to the state level and to decadal (vs annual) analysis; the IV estimate of AHA hospital expenditures in this specification is generally similar in magnitude although somewhat less precise than that in our baseline specification.

[44]We also examined the elasticity of various components of state-level health care utilization from the NHIS. The NHIS data cover 1973-1990 (data before 1973 do not have state identifiers) and are not interpolated, which is a clear advantage relative to the HCFA data. On the other hand, the NHIS only measures utilization on the extensive margin. This implies that NHIS data will not be informative about increases in expenditure on the intensive margin. As in the AHA data, we find no evidence in the NHIS of a positive income elasticity of hospital utilization. We also find no evidence of a positive income elasticity of doctor visits (indeed, the point estimates are negative, though not statistically significant). Results available on request.

that the income elasticity of overall health expenditures is unlikely to be significantly higher than the income elasticity of hospital spending. This is also consistent with the time-series evidence in Figure 1. We therefore conclude that our estimates of the income elasticity of hospital spending are likely to be representative of the income elasticity of total health expenditures.

2.3.3 Labor Income Versus Total Income

Another potential concern with our main data is that our baseline income measure captures only the effect of our instrument on labor income. If capital income and labor income do not respond proportionately to our instrument, we may be under-stating (or over-stating) the first-stage relationship, and consequently, over-stating (or under-stating) the income elasticity in the second stage. Unfortunately, annual data on labor and capital income do not exist for our time period at a level of disaggregation below the state.

We therefore investigate how our estimates at the state level change when we use Gross State Product (GSP) as our measure of income, rather than our baseline payroll measure; unlike payroll, GSP includes both labor and capital income. Table 7 shows the results of this exercise. Panel A shows the IV estimates, and Panel B shows the first-stage estimates. Columns 1 and 2 compare results at the state level when labor (payroll) income and GSP are used, respectively, as our income measure. The first stage suggests that, in response to our instrument, non-labor income appears to rise by the same proportion, or by slightly more, then our primary measure of labor income (compare columns 1 and 2 of Panel B). If anything, therefore, the results suggest that the estimates using labor income only may be slightly over-stating the income elasticity of health expenditures (compare columns 1 and 2 of Panel A).

Since, as discussed, we lose variation by aggregating to the state level, we also report results at the state level when we include the entire US in the sample rather than just the 16 states in the South. Column 3 shows the results when we use labor income (from the CBP payroll data) as our measure of income and column 4 shows the results when we use the GSP measure, which incorporates capital income. Once again the results suggest that non-labor income may rise slightly more than proportionately with labor income, so that our income elasticities in our baseline estimates may be slightly overstated.[45]

2.3.4 Heterogeneity in Income Elasticities

Another potential concern with our conclusions concerning the role of rising incomes in explaining the rising health share of GDP is that our IV estimates are based on a specific type of income variation as well as a specific area of the country and time period. If there is substantial heterogeneity in the income elasticity of health expenditures across any of these dimensions, out-of-sample extrapolations may be particularly unreliable. We therefore explored whether there appears to be substantial heterogeneity in our estimated income elasticity. All in all, we read the available evidence as suggesting that the quantitative estimates are reasonably similar across different sources of income variation, geographic samples, time periods, and time horizons; we therefore do not see any reason to suspect that heterogeneous elasticities are likely to lead to a serious underestimation of the effect of rising incomes on health care expenditures.

Source and extent of income variation At a general level, one might be concerned that the source and range of the variation in income that we are exploiting may be insufficient to estimate (or detect) income elasticities significantly greater than one. To alleviate this concern, we estimated similar IV regressions with spending on goods that can be classified as a luxury on a priori grounds (e.g., recreation). Since we do not have data on spending on other goods at the ESR level, we pursued this strategy at the state level using data on industry-specific Gross State Products (GSP) for other service industries. Specifically, we used our instrument at the state level to examine the income elasticity of four potential luxury goods: "amusement and recreation services," "hotels and other lodging places," "legal services" and "other

[45] The results in column 3 also suggest that our estimates are not sensitive to using the entire United States. In later robustness analysis we show this is true at the ESR level as well (see Table 10 below).

services," which includes (among other things) record production, actuarial consulting, music publishing, and other consulting.[46] We also estimated the income elasticity of "food and kindred products," which we expect to be a necessity. The results are shown in Table 8.[47]

The results suggest that our source of variation in income is strong enough to uncover elasticities greater than one at the state level.[48] Legal services and "other services" both appear to be strong luxuries. Amusement services and hotels also show an income elasticity of close to or above 1. By contrast, food stores appear to be a necessity, with an income elasticity that is virtually the same as what we estimate for health services (see column 6).

A more specific concern is that, as discussed in Section 2.2.1, we cannot reject that our income variation at the ESR level comes entirely from changes in employment at roughly constant wages (see Table 3), while about half of income growth in the United States over the last half century comes from increased wages per employed individual (US Census Bureau, 2008).[49] This raises the potential concern that, if the elasticity of health spending with respect to income is increasing in income, the elasticity of health care spending with respect to increases in wages may be larger than the elasticity with respect to increases in employment.

Table 9 investigates whether there is any evidence of this type of convexity in Engel curves for health expenditures. Column 1 reports results from the baseline IV specification, while column 2 adds an interaction of the ESR's (log) income with its (log) income in 1970. This strategy allows the effect of changes in income to vary based on initial income levels and provides a simple check against the possibility that

[46] A complete definition of "other services" can be found here: `http://www.osha.gov/pls/imis/sic_manual.display?id=1014&tab=description`.

[47] An estimate for health services GSP, which was already discussed in subsection 2.3.2, is also included in this table.

[48] More information on each of these categories can be found here: `http://www.bea.gov/regional/gsp/default.cfm?series=SIC`. First-stage results for this same specification are shown in Table 7, Panel B, columns 1 and 3. Second stage results for this same specification using the AHA hospital expenditure data as the dependent variable can be found in Table 7, Panel A, columns 1 and 3.

[49] At the state level we estimate that our instrument is associated with a statistically significant increase in wages, although the increase in income is still predominantly due to an increase in employment (not shown).

the income elasticity of health expenditures may vary systematically with the level of income of the area. We instrument for log income and the interaction of log income with 1970 ESR log income with our standard instrument (oil reserves times log oil prices) and the interaction of this instrument with 1970 ESR log income. The results show no evidence that the Engel curve for health expenditures is convex; if anything the point estimates suggest a (statistically insignificant) concave Engel curve.

As another check on the potential convexity of the relationship between income and hospital spending, we looked for nonlinearities in the reduced-form relationship. Column 3 reproduces the baseline reduced-form results for comparison and column 4 reports the results of a modified reduced-form specification, which also includes the square of the baseline instrument (i.e., $(\log p_{t-1} \times I_j)^2$ as well as $\log p_{t-1} \times I_j$). The estimates in column 4 also show no evidence of a convex relationship between income and health expenditures. The lack of any convexity in the relationship between income and health spending further suggests that the income elasticity of health expenditures is unlikely to be significantly greater at higher levels of income or for larger income changes.

Finally, we note that because oil prices both rise and fall over our time period, our instrument predicts both increases and decreases in income. From a purely estimation standpoint, this is a strength of our instrument, since it makes it less likely that it simply captures differential (monotonic) trends across different areas of the country. Nevertheless, since much of the motivation of our paper is related to the effects of rising incomes on health care expenditures, we also investigated whether the effects of rises and declines in income are asymmetric. In particular, we re-estimated our baseline models allowing positive and negative changes (between t and $t-1$) in income to have different effects (and we instrumented these income variables with our baseline instrument interacted with an indicator for whether oil prices rose between dates t and $t-1$). We found no evidence of such asymmetric effects (results available upon request).

Different areas and time period Table 10 explores the sensitivity of our estimates to defining the sample based on different geographic regions and different time periods. Panel A shows the IV estimates and Panel B shows the corresponding first-stage results. Column 1 reproduces our baseline estimates, which are for the 16 Southern states focusing on the time period 1970-1990.

As discussed above, we chose to limit our baseline sample to the Southern United States both because the oil reserves are concentrated in the South and because the ESRs in this region are more comparable, thus less likely to experience differential trends in hospital spending owing to other reasons. In column 2 we further limit the sample to the 7 Southern states that have oil reserves in our data. The results are quite similar. In column 3 we go in the opposite direction, and look at the entire United States. The results in this column show that expanding the sample to the entire United States (not including Alaska and Virginia) results in a very similar point estimate of the income elasticity (0.804 vs. 0.723 in the baseline), though the estimate is less precise (standard error = 0.631 compared to 0.214 in the baseline).[50]

We also explored whether within the South our estimates were sensitive to excluding a particular state. Appendix Table A2 shows the results from estimating our baseline specification (from column 1) dropping each one of the 16 states at a time. The results indicate that the estimates are generally quite robust both in terms of magnitude and precision to the omission of a single state. The exception occurs when we exclude Texas. In this case, the point estimate falls by about 40 percent; combined with the increase in standard error, this makes the estimate of the income elasticity of hospital expenditure no longer significant at the 5% level. This is not surprising since much of the variation in oil intensity in our sample is within Texas (see Figure 3).

Our baseline time period is for 1970-1990 and covers the original oil boom and bust. In column 4 of Table 10, we return to our baseline Southern states sample, but now expand the time period 1970-2005 (thus including all available years with

[50]We do not include Alaska because of the Alaska Permanent Fund (established in 1976), as well as the difficulty in forming consistent data by ESR between 1970 and 1990. We do not include Virginia because of the difficulty in forming consistent data by ESR between 1970 and 1990.

data). Figure 2 shows that oil prices experienced a second boom starting in 1999. Nevertheless, we lose the first stage when we include the post 1990 years (and therefore do not report the corresponding IV estimate). This weaker first-stage relationship appears to reflect the inadequacy of imposing constant ESR fixed effects over a 36 year period. Indeed, when this assumption is relaxed in column 5 by including state-specific time trends, the first-stage relationship is again statistically significant and leads to an IV estimate of similar magnitude to the baseline.

Permanent versus transitory elasticities The interpretation of our estimates depend on whether oil price changes are permanent or transitory. This is investigated in Table 11 using the time-series data shown in Figure 2. Column 1 shows that a regression of the log oil price at time t on its one year lag produces a coefficient of 1.009 (standard error = 0.043). The augmented Dickey-Fuller unit-root test reported at the bottom comfortably fails to reject the null hypothesis that log oil prices follow a unit root. The remaining columns of this table show several different specifications, all indicating that we cannot reject that changes in oil prices are permanent. These findings are consistent with those of previous researchers.[51] The available evidence therefore suggests that our empirical strategy speaks to the effects of permanent (rather than transitory) changes in income on health care expenditures.

Short-run versus long-run income elasticities Since we focus on annual variation, our empirical strategy estimates the short-run response of health expenditures to (permanent changes in) income. This may naturally be different from the long-run response of health expenditures. For example, increased demand may result in the short run in higher prices, with the response of quantities emerging with a delay as capacity expands. However, there are no strong theoretical reasons to expect the long-run income elasticity to be greater than the short-run elasticity. For example, if health care demand is inelastic (with price elasticity less than one, which is plausible,

[51] Kline (2008) conducts a more detailed analysis of the time-series behavior of oil prices and concludes that oil prices are "well approximated by a pure random walk". See also Hamilton (2008) for a similar conclusion.

for example, because of insurance), as capacity expands in the long run in the face of rising incomes, overall health expenditures will increase less than in the short run. In addition, if long-run increases in income also improve overall health, the long-run increase in health expenditures may again be less than in the short run. Nevertheless, even though there are no a priori reasons to expect long-run effects to be greater than short-run effects, it is important to understand whether our empirical strategy is estimating the former or the latter.

To investigate this issue, we re-estimated our regressions using decadal observations, thus removing the source of variation due to short-run changes in our instrument. Table 12 compares our baseline results—which use annual observations from 1970-1990 in columns 1 through 3—with the estimates using only decadal observations (1970, 1980, 1990) in columns 4 through 6. With only the decadal observations, the first stage is only slightly weaker (compare columns 4 and 1). The IV elasticity estimate from the decadal estimate is similar to the baseline annual estimate (0.794 compared to 0.723) although the standard error of the decadal estimate is roughly double what we obtain with annual data. We read these results as suggestive of a long-run income elasticity that is similar to the short-run elasticity.

This conclusion also receives support from the lack of capacity responses. If long-run effects were significantly larger than short-run effects, we would expect to see hospitals expanding capacity (either simultaneously with the increase in health expenditures or gradually as they reach their capacity constraints). However, Table 5 showed no evidence of an increase in hospital capacity or utilization (in particular, there was no increase in admissions, patient days, hospital beds, and hospital entry in response to the rise in local income).

A related issue is that there might be heterogeneity in the adjustment dynamics of hospital spending in response to increases in income. For example, suppose that some of the ESRs respond immediately to increases in income, while other ESRs take one or two years to respond. In this case, results using the annual panel and assuming immediate and complete adjustment would underestimate the true long-run income elasticity. We show in the on-line Appendix (see section C) that specifications using

3-year averages typically perform better when there are heterogeneous adjustment dynamics by ESR. Thus in column 7 we report results based on 3-year averages. The estimated elasticity increases slightly (from 0.723 to 0.826).

2.3.5 Robustness

We also performed a large number of robustness checks of our baseline estimates, designed to explore the robustness of our instrumental-variables estimates along a number of dimensions and to examine the validity of our identifying assumption. Specifically, we explored a variety of alternative specifications designed to investigate the validity of our identifying assumption; we examined the robustness of our results to alternative specifications of our instrument; and we explored alternative ways to address potential serial correlation in the residuals. The results from these additional analyses were in general quite reassuring. They are presented in detail in the on-line Appendix (Section B and Tables A3-A6).

2.4 Conclusion

This paper has explored the role of the secular rise in incomes in the dramatic run-up in the health share of GDP in the United States, which increased from 5 percent of GDP in 1960 to 16 percent in 2005. A common conjecture is that rising incomes have played a primary role in the increase in the health share of GDP. A finding of a primary role for rising incomes would have important implications for forecasting the future growth of the health share of GDP. It would also provide crucial input into an investigation of the potential optimality (or sub-optimality) of rising health share of GDP. Yet, surprisingly, little is known about the empirical impact of rising aggregate incomes on health spending.

We attempted to estimate the causal effect of aggregate income on aggregate health expenditures by instrumenting for local area income with time-series variation in global oil prices interacted with cross-sectional variation in the oil reserves in different areas of the Southern United States. This strategy is attractive not only because it isolates a potentially-exogenous source of variation in incomes but also because it incorporates local general equilibrium effects, as we estimate the response of health expenditures in the area to an aggregate change in incomes. We also presented evidence suggesting that national or global general equilibrium effects of rising income on health expenditures – which our estimates would not capture – are unlikely to be quantitatively important.

Across a wide range of specifications, we estimate a positive and statistically significant income elasticity of hospital expenditures that is almost always less than 1. Our central estimate is an income elasticity of 0.72 (standard error = 0.21). This estimate is reasonably robust to a range of alternative specifications.

Our central point estimate suggests that rising income did not contribute to the rise in the health *share* of GDP between 1960 and 2005. Our 95 percent confidence interval—which includes at its upper end an income elasticity of 1.1—suggests that we can reject a role of rising income of explaining more than a very small part, 0.5 percentage points, of the 11 percentage point increase in the health share of GDP over that time period. Although considerable caution is warranted in extrapolating estimates from a particular source of variation, time period, and part of the country to the overall impact of rising incomes in the post-war period, we provided additional evidence suggesting that many of the most salient potential concerns with such extrapolation are not likely to pose major threats to our conclusions.

While our findings suggest that the increase in income is unlikely to be a primary driver of the increase in the health share of GDP, they do not provide an answer to the question of what is behind this notable trend. There is general consensus that rapid progress in medical technologies is a (or "the") major driver of increasing health expenditures (e.g., Newhouse, 1992, Fuchs, 1996, Cutler, 2002, Congressional Budget Office, 2008), though presumably technological progress itself is being spurred by other factors. Our analysis thus indirectly also suggests that rising incomes are unlikely to be the major driver of medical innovations either. An interesting possibility is that institutional factors, such as the spread of insurance coverage, have not only directly encouraged increased spending but also induced the adoption and diffusion of new medical technologies (Weisbrod 1991, Finkelstein 2004, Finkelstein 2007, Acemoglu and Finkelstein, 2008). This channel of induced innovation could not only account for the increase in the health share of GDP in the United States, but provided that technological advances in the United States spread relatively rapidly to other advanced economies, it could also be a major contributor to the similar trends experienced by other OECD countries. An investigation of this possibility, as well as more general analyses of the determinants of technological change in the health care sector, are important and interesting areas for further work.

References

Abraham, Katherine and John C. Haltiwanger. 1995. "Real Wages and the Business Cycle." *Journal of Economic Literature* 33(3): 1215-1264.

Acemoglu, Daron. 2002. "Directed Technical Change." *Review of Economic Studies* 69(4).

Acemoglu, Daron. 2007. "Equilibrium Bias of Technology." *Econometrica* 75(5).

Acemoglu, Daron and Amy Finkelstein. 2008. "Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector." *Journal of Political Economy*, 116(5): 837-880 .

Acemoglu, Daron and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics*, 119, 1049-1090.

Baker, Laurence. 2001. "Managed Care and Technology Adoption in Health Care: Evidence from Magnetic Resonance Imaging." *Journal of Health Economics*, 20: 395-421.

Baker, Laurence and Ciaran Phibbs. 2002. "Managed care, technology adoption, and health care: the adoption of neonatal intensive care." *Rand Journal of Economics* 33: 524-548.

Black, Dan, Terra McKinnish and Seth Sanders. 2005. "The Economic Impact of the Coal Boom and Bust." *Economic Journal*, 115(2): 449-476.

Blanchard, Olivier J. and Larry F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity.*

Blomqvist, A.G. and Carter R.A.L.. 1997. "Is Health Care Really a Luxury?" *Journal of Health Economics*, 16(2): 207-229.

Buckley, Patrick D. 2003. Essays in Applied Microeconomics: Signaling, Marriage Hazard Rates, and Ability Sorting. Ph.D. thesis, MIT, Department of Economics.

CMS. 2006. National Health Expenditures Data, `http://www.cms.hhs.gov`.

Congressional Budget Office. 2008. Technological Change and the Growth of Health Spending, `http://www.cbo.gov/ftpdocs/89xx/doc8947/01-31-TechHealth.`

pdf.

Cutler. 1995. "Technology, Health Costs, and the NIH." National Institutes of Health Roundtable on the Economics of Biomedical Research.

Cutler, David. 2002. "Equality, Efficiency, and Market Fundamentals: The Dynamics of International Medical Care Reform." *Journal of Economic Literature*, 881-906.

Cutler, David and Ellen Meara. 2003. "Changes in the Age Distribution of Mortality over the 20th Century." In David Wise (ed) Perspectives on the Economics of Aging. Chicago: University of Chicago Press.

Cutler, David and Louise Sheiner. 1998. Managed Care and the Growth of Medical Expenditures." In Alan Garber (ed). *Frontiers in Health Policy Research.*

Davis, Steven and John Haltiwanger. 2001. "Sectoral job creation and destruction responses to oil price changes." *Journal of Monetary Economics*, 48: 465-512.

Finkelstein, Amy. 2004. "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry." *Quarterly Journal of Economics*, 527-564.

Finkelstein, Amy. 2007. "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare." *Quarterly Journal of Economics*, 122(3): 1-37.

Fogel, Robert. 1999. "Catching up with the Economy." *American Economic Review*, 89(1): 1–21.

Fuchs, Victor. 1996. "Economics, Values, and Health Care Reform." *American Economic Review*, 86(1): 1-24.

Gerdtham, Ulf, and Bengt Jonsson. 2000. "International Comparisons of Health Expenditure: theory, data and econometric analysis" in AJ Culyer and J Newhouse (eds). Handbook of Health Economics. Amsterdam: Elsevier.

Gruber, Jonathan. 2003. "Medicaid", in Robert Moffitt (ed) Means Tested Transfer Programs in the U.S. Chicago: University of Chicago Press: 15 - 77.

Hall, Robert and Charles Jones. 2007. "The Value of Life and the Rise in Health Spending." *Quarterly Journal of Economics.* February.

Hamilton, James. 2008. "Understanding Crude Oil Prices." NBER Working Paper No. 14492.

Kilian, Lutz. 2008. "The Economic Effects of Energy Price Shocks." *Journal of Economic Literature*, 46(4): 871-909.

Kline, Patrick. 2008. "Understanding Sectoral Labor Market Dynamics: An Equilibrium Analysis of the Oil and Gas Field Services Industry," <`http://www.econ.yale.edu/~pk279/papers/oil.pdf`>.

Levit, Katherine. 1982. "Personal Health Care Expenditures by State, selected years, 1966-78." *Health Care Financing Review*, 4(2).

Levit, Katherine. 1985. "Personal Health Care Expenditures by State: 1966-82." *Health Care Financing Review*, 6(4).

Michaels, Guy. 2007. "The Long-Term Consequences of Resource Based Specialization." London School of Economics, Mimeo.

Moran, John and Kosali Simon. 2006. "Income and the Use of Prescription Drugs by the Elderly." *Journal of Human Resources*, XLI(2): 411-432.

Murphy, Kevin M. and Robert H. Topel. 2003. "The Economic Value of Medical Research", in Measuring the Gains from Medical Research: An Economic Approach, edited by Kevin M. Murphy and Robert H. Topel. Chicago: University of Chicago Press.

Murphy, Kevin M. and Robert H. Topel. 2006. "The Value of Health and Longevity" *Journal of Political Economy*, 114(5): 871-904.

Newey, Whitney K., James L. Powell and Francis Vella. "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67(3): 565-603

Newhouse, Joseph and the Insurance Experiment Group. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment.* Harvard University Press: Cambridge MA.

Newhouse, Joseph. 1992. "Medical Care Costs: How Much Welfare Loss"? *Journal of Economic Perspectives*, 6(3): 3-21.

Newhouse, Joesph. 1977. "Medical Care Expenditures: A Cross-national survey." *Journal of Human Resources*, 12 (1): 115-125.

Newhouse, Joseph and Charles Phelps. 1976. "New Estimates of Price and Income Elasticities for Medical Care Services." In Richard Rosett (ed), The Impact of Health

Insurance on the Health Services Sector. New York: National Bureau of Economics Research.

Nordhaus, William D. 2003. "The Health of Nationas: The Contributon of Improved Health to Living Standards" in Measuring the Gains from Medical Research: An Economic Approach, edited by Kevin M. Murphy and Robert H. Topel. Chicago: University of Chicago Press.

OECD 2004. OECD Health Data 2000. <`http://www.oecd.org/document/16/0,2340,en_2649_34631_2085200_1_1_1_1,00.html`>.

OECD. 2006. "Projecting OECD Health and Long-Term Care expenditures: What are the drivers?" OECD Economics Department working paper #477. <`http://www.oecd.org/dataoecd/57/7/36085940.pdf`>.

Oil and Gas Journal Data Book. 2000. Tulsa, Oklahoma: PennWell Books.

Siegal, Jacob S. 2002. "Applied Demography: Applications to Business, Government, Law and Public Policy." Emerald Group Publishing.

Solon, Gary, Robert Barsky and Jonathan Parker. 1994. "Measuring the Cyclicality of Real Wages: How Important is the Composition Bias?" *Quarterly Journal of Economics*, 109(1): 1-25.

US Census Bureau. Various States and Years. Current Population Reports, Series 25, Population Estimates and Projections.

US Census Bureau, Statistical Abstract of the United States: 2008 (127th Edition) Washington, DC, 2007; <`http://www.census.gov/compendia/statab`>.

Weisbrod, Burton. 1991. "The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment." *Journal of Economic Literature*, 29(2): 523-552.

Table 1: Descriptive Statistics

Variable	ESR-year data Mean	ESR-year data Standard Deviation	State-year data Mean	State-year data Standard Deviation
Oil and Gas Data Book Data				
Oil Reserves (million barrels)	532.3	1596.1	3371.7	9124.3
County Business Patterns Data				
Total Income (Payroll); ($millions)	2916.9	6066.7	18494.4	20751.0
Total Employment (millions)	0.21	0.35	1.32	1.13
AHA Hospital Data				
Total Expenditures ($millions)	292.61	636.31	1854.22	2257.94
Hospital Payroll ($millions)	139.87	284.12	886.40	1011.38
Admissions (millions)	0.11	0.16	0.73	0.55
Inpatient Days (millions)	1.08	1.47	6.85	4.76
Beds (thousands)	4.15	5.65	26.29	19.06
Full-time Equivalents (thousands)	9.58	14.55	60.72	48.58
RN / (LPN + RN)	0.63	0.12	0.69	0.09
# of Technologies	46.98	18.10	48.37	19.53
# of Hospitals	24.67	26.57	156.43	126.71
Current Population Reports and NHIS Data				
Population (millions)	0.68	0.89	4.44	3.40
HUWP (millions)	0.60	0.84	3.88	3.02
BEA GSP Data (all in $millions)				
Total GSP			54559.5	60731.7
(Industry-Specific GSPs)				
Health Services			1639.9	2182.0
Amusement and Recreation Services			150.3	266.4
Hotels and Other Lodging			237.7	343.6
Legal Services			312.9	575.2
Other Services			624.5	995.6
Food			524.3	485.0
Health Care Financing Administration (HCFA) Data (all in $millions)				
Total Health Care Expenditures			5923.8	6447.2
Hospital Expenditures			2641.1	2663.7
Physician and Other Services			1626.0	2065.4
Dental Services			303.3	330.7
Drugs and Other Medical Non-durables			654.7	685.4
Vision Products			106.2	130.8
Nursing Care			390.7	382.2
Other Health Services			201.8	425.2

Notes: Summary statistics in columns 1 and 2 are for the baseline sample of 99 economic sub-regions (ESRs) in the 16 Southern states between 1970 and 1990 (i.e. all statistics are ESR-year); columns 3 and 4 report summary statistics for the State-year data for the same baseline sample of 16 southern states between 1970 and 1990. Source for variables is given in italics. BEA and HCFA data are only available at state level. N = 2065 at ESR-year except for RN/(LPN+RN) which is 1576 and Inpatient Days which is 1967. N = 326 at State-year except for HCFA data and except for RN/(LPN+RN) which is 251 and Inpatient Days which is 311. Data on RNs and LPNs are only available in 1970, 1972, 1974, 1976, 1978, and 1980-1990. Data on Inpatient Days are not available in 1979. N = 236 at State-year for HCFA data which are only available in 1972, 1976-1978, and 1980-1990. HUWP is a hospital-utilization weighted measure of population. See text for more details.

Table 2: Comparing economic sub-regions in 1970 With Different Oil Reserves

Variable	(1) Mean for ESRs with Large Oil Wells	(2) Mean for ESRs without Large Oil Wells	(3) Coefficient	(4) p-value
Population (in millions)	0.687	0.521	0.113	0.155
Total Employment (in millions)	0.168	0.137	0.075	0.306
Hospital Expenditures (in $thousands)	0.059	0.050	0.072	0.356
Hospital Beds (in thousands)	4.671	3.940	0.094	0.184
Total Income (in $thousands)	0.989	0.778	0.077	0.298
p-value of F-test of joint significance				0.357
(F-statistic = 1.12 for $F(5,92)$)				

Notes: All results based on 1970 cross-section of the ESRs in the baseline sample (i.e. the 16 Southern states). Column 3 reports the coeffient from a regression of Oil Reserves on the variable in the row header and a constant term; in these regressions in column 3, both dependent and independent variables are standardized to have standard deviation of 1. Column 4 reports the associated p-value (based on heteroskedasticity-robust standard errors). The final row of table reports results from a regression of Oil Reserves on all of the variables listed in table and a constant term. N = 98 in the regressions reported in columns 3 and 4 because AHA data for Washington, DC are not available in 1970. N = 30 in column 1 and 68 in column 2.

Table 3: First Stage

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Geographic level of analysis:	economic sub-region							State
Dependent variable:	Total Income	Total Employment	Population	Income per capita	Population < 55	Population $\geq$ 55	Income per HUWP	Income per HUWP
Oil Reserves$_j$ $\times$	12.900	15.542	5.252	7.648	6.421	1.545	9.245	2.564
log(oil price)$_{t-1}$	(2.980)	(2.572)	(1.491)	(1.937)	(1.756)	(1.531)	(2.271)	(0.523)
	[0.001]	[0.000]	[0.003]	[0.001]	[0.002]	[0.329]	[0.001]	[0.000]
R^2	0.994	0.969	0.997	0.984	0.997	0.996	0.983	0.989
N	2065	2065	2065	2065	2065	2065	2065	326
F-statistic	18.74	36.53	12.40	15.58	13.37	1.02	16.58	24.05

Notes: Table reports results from estimating variants of equation (8) and (11) by OLS. Dependent variables are defined in column headings and are all in logs; in column 7 and 8 the dependent variable is income divided by a hospital-utilization weighted measure of population (HUWP). The sample is all Southern states between 1970 and 1990. Unit of observation is an economic sub-region (ESR)-year except in column 8 where it is State-year. All models include ESR (or state in column 8) and year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets.

Table 4: Hospital Expenditures

	(1)	(2)	(3)	(4)	(5)	(6)
Geographic level of analysis:	economic sub-region					State
Population adjustment:	HUWP	HUWP Reduced Form	HUWP	None	Per Capita	HUWP
	OLS	OLS	IV	IV	IV	IV
$\log(\text{Income})_{jt}$	-0.027		0.723	0.801	0.665	0.550
	(0.074)		(0.214)	(0.155)	(0.263)	(0.230)
	[0.723]		[0.004]	[0.000]	[0.023]	[0.030]
Oil Reserves_j × $\log(\text{oil price})_{t-1}$		6.680				
		(2.048)				
		[0.005]				
R^2	0.973	0.973	0.968	0.989	0.970	0.992
N	2065	2065	2065	2065	2065	326

Notes: Table reports results of estimating equations (7), (10) or (12) by OLS or IV as indicated. Dependent variable is log hospital expenditures. In columns 1, 2, 3, and 6, both hospital expenditures and income are divided by a hospital-utilization weighted measure of population (HUWP) before taking logs (see equations (10) through (12)). In column 4 hospital expenditures and income are not adjusted before taking logs, and in column 5 both hospital expenditures and income are divided by the total population before taking logs. The sample is all Southern states between 1970 and 1990. Unit of observation is an economic sub-region (ESR)-year except in column 6 where it is a state-year. All models include ESR (or state in column 6) and year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets.

Table 5: Other Hospital Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent Variable:	Total Hospital Expenditures	Total Hospital Payroll	FTE	RN/ (RN+LPN)	Admissions	In-Patient Days	Beds	# of Hospitals	# of Technologies	Open-Heart Surgery	Radioisotope Therapy
$\log(\text{Income})_{jt}$	0.723	0.934	0.039	0.329	-0.430	-1.034	-0.698	-0.552	-0.132	-3.163	1.083
	(0.214)	(0.233)	(0.222)	(0.089)	(0.193)	(0.488)	(0.455)	(0.358)	(0.221)	(11.334)	(2.575)
	[0.004]	[0.001]	[0.862]	[0.002]	[0.042]	[0.051]	[0.146]	[0.144]	[0.558]	[0.169]	[0.545]
R^2	0.968	0.958	0.893	0.868	0.788	0.884	0.871	0.981	0.945		
N	2065	2064	2065	1576	2065	1967	2065	2065	2065	849	262

Notes: Columns 1 through 9 report IV estimates of equation (10) with the first stage given by equation (11). Column 1 reproduces baseline results from column 3 in Table 4. Unit of observation is an economic sub-region (ESR)-year. The baseline sample is all Southern states between 1970 and 1990. Each column shows results for a different dependent variable, as indicated in the column heading. Dependent variables in columns 1-3 and 5-7 are in logs and are divided (before taking logs) by a hospital-utilization weighted measure of population (HUWP). Dependent variables in columns 8 and 9 are in logs but not adjusted by any population measure; dependent variable in column 4 is not adjusted by any population measure and is not in logs. Columns 10 and 11 report results from an instrumental variables estimator of the Cox proportional hazard model shown in equation (13). Dependent variable in columns 10 and 11 is an indicator variable for whether an at-risk ESR adopts the technology in that year and sample size reflects the number of ESRs "at risk" for adoption in each year. In column 10, there are 56 ESRs that have not adopted open-heart surgery technology by 1970 and 22 ESRs that have not adopted by 1990. In column 11, there are 21 ESRs that have not adopted radioisotope therapy by 1972 (the first year data are available) and 8 ESRs that have not adopted by 1990. Data for RNs and LPNs (column 4) only exist in 1970, 1972, 1974, 1976, 1978, and 1980-1990. Data for in-patient days (column 6) do not exist in 1979. All models include ESR and year fixed effects, except columns 10 and 11 which have region fixed effects and controls for total hospital beds and hospital expenditures in 1970. In all columns income is divided by HUWP before taking logs. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets; in columns 10 and 11 the standard errors and p-values are bootstrapped (clustered by state).

Table 6: Hospital Spending Versus Overall Health Spending

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Regression:	First Stage OLS	IV	IV	IV	IV	IV	IV	IV	IV
Dependent Variable:	Income	Total Health Care Exp.	Hospital Exp.	Physician and Other Services	Dental Services	Drugs and Other Medical Non-durables	Vision Products	Nursing Care	Other Health Services
Panel A: Southern States Only									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Oil Reserves$_j$ × log(oil price)$_{t-1}$	3.626								
	(0.776)								
	[0.000]								
log(Income)$_{jt}$		0.055	0.067	0.179	0.622	0.248	1.187	-1.302	-0.359
		(0.077)	(0.157)	(0.152)	(0.100)	(0.120)	(0.516)	(0.321)	(0.228)
		[0.484]	[0.675]	[0.257]	[0.000]	[0.057]	[0.036]	[0.001]	[0.137]
R^2	0.985	0.998	0.995	0.996	0.991	0.993	0.914	0.926	0.963
N	236	236	236	236	236	236	236	236	236
F-statistic	21.81								
Share of Total Health Care Exp.			46.30%	24.73%	5.17%	11.33%	1.80%	7.02%	3.44%
Panel B: All U.S.									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Oil Reserves$_j$ × log(oil price)$_{t-1}$	3.162								
	(0.586)								
	[0.000]								
log(Income)$_{jt}$		0.098	0.139	0.365	0.650	0.307	0.748	-1.944	-0.953
		(0.167)	(0.151)	(0.186)	(0.173)	(0.112)	(0.824)	(0.968)	(0.758)
		[0.558]	[0.361]	[0.056]	[0.000]	[0.009]	[0.368]	[0.050]	[0.214]
R^2	0.98	0.996	0.965	0.974	0.986	0.989	0.879	0.918	0.915
N	729	729	729	729	729	729	729	729	729
F-statistic	29.11								
Share of Total Health Care Exp.			45.06%	25.04%	6.07%	10.40%	2.02%	8.57%	3.39%

Notes: Table reports first stage results of estimating equation (11) by OLS in column 1; remaining columns report estimates of variants of estimating equation (10) by IV. Unit of observation is a State-year in all columns. Dependent variables are various measures of health care expenditures from the Health Care Finance Administration (HCFA). HCFA data are available in 1972, 1976 - 1978, and 1980-1990. All dependent variables and income are in logs and divided by a hospital-utilization weighted measure of population (HUWP). In all columns income is divided by HUWP before taking logs. Sample is Southern states in Panel A and All U.S. (except Alaska and Virginia) in Panel B. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets.

Table 7: Labor Income vs. All Income

Panel A: IV Results

Dependent Variable: Hospital Expenditures

	(1)	(2)	(3)	(4)
$\log(\text{Income})_{jt}$	0.550	0.451	0.740	0.568
	(0.230)	(0.160)	(0.359)	(0.263)
	[0.030]	[0.013]	[0.045]	[0.036]
R^2	0.992	0.993	0.981	0.982
N	326	326	1015	1015

Panel B: First Stage Results

Dependent Variable: Income

	(1)	(2)	(3)	(4)
Oil Reserves_j × $\log(\text{oil price})_{t-1}$	2.564	3.128	2.220	2.895
	(0.523)	(0.851)	(0.443)	(0.682)
	[0.000]	[0.002]	[0.000]	[0.000]
R^2	0.989	0.990	0.985	0.983
N	326	326	1015	1015
F-statistic	24.05	13.50	25.10	18.05

Specification

	(1)	(2)	(3)	(4)
Income definition	Payroll	GSP	Payroll	GSP
Geographic sample	South	South	USA	USA

Notes: Table reports estimates of variants of estimating equation (10) by IV in Panel A and equation (11) by OLS in Panel B. Unit of observation is a State-year in all columns. In all specifications income and hospital expenditures are divided by hospital-utilization weighted measure of population (HUWP) and then logged. Bottom rows define the specification variants; these are the definition of income (Payroll as in the baseline specification or Gross State Product (GSP)) and the geographic sample (South or all US). In all columns the years of analysis are 1970 - 1990. The sample is all Southern states between 1970 and 1990 in columns 1 and 2; columns 3 and 4 expand sample to all US (except Alaska and Virginia). Column 1 reproduces results from column 6 in Table 4 (Panel A) and column 8 of Table 3 (Panel B). All regressions include state and year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets.

Table 8: Income Elasticity of Other Goods

	(1)	(2)	(3)	(4)	(5)	(6)
	Industry-specific Gross State Product					
	Amuse-ment	Hotels	Legal Services	Other Services	Food	Health Services
	Panel A: Southern States Only					
$\log(\text{Income})_{jt}$	0.900	0.835	1.635	1.375	-0.009	-0.048
	(0.385)	(0.319)	(0.317)	(0.387)	(0.416)	(0.181)
	[0.034]	[0.019]	[0.000]	[0.003]	[0.984]	[0.793]
R^2	0.984	0.984	0.991	0.989	0.965	0.996
N	326	326	326	308	324	326
	Panel B: All U.S.					
$\log(\text{Income})_{jt}$	1.080	0.940	1.749	1.400	0.255	0.207
	(0.384)	(0.397)	(0.291)	(0.270)	(0.356)	(0.412)
	[0.007]	[0.022]	[0.000]	[0.000]	[0.477]	[0.617]
R^2	0.975	0.978	0.988	0.984	0.977	0.994
N	1013	1015	1015	989	1013	1015

Notes: Table reports results from estimating variants of equation (10) by IV. Dependent variables are given in column headings. All dependent variables are in logs, and all dependent variables and income are divided by a hospital-utilization weighted measure of population (HUWP) before taking logs. The sample is all Southern states between 1970 and 1990 in Panel A and all US states (except Alaska and Virginia) between 1970 and 1990 in Panel B. Unit of analysis is a state-year. All columns include state and year fixed effects. Dependent variable is the Gross State Product for various industries, as indicated by column headings. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets.

Table 9: Decomposition and Tests for Nonlinear Effects

Dependent Variable: Hospital Expenditures	(1)	(2)	(3)	(4)
			Reduced Form	Reduced Form
Regression:	IV	IV	OLS	OLS
Oil Reserves$_j$ × log(oil price)$_{t-1}$			6.680	10.567
			(2.099)	(7.511)
			[0.006]	[0.180]
log(Income)$_{jt}$	0.725	0.833		
	(0.216)	(0.369)		
	[0.005]	[0.040]		
log(Income)$_{jt}$ × log(Income)$_{j,t=1970}$		-0.066		
		(0.143)		
		[0.652]		
{ Oil Reserves$_j$ × log(oil price)$_{t-1}$ }^{2}				-487.728
				(717.177)
				[0.507]
R^2	0.967	0.965	0.973	0.973
N	2054	2054	2065	2065
-1 standard deviation	0.725	0.862	6.680	11.855
Marginal Effect at Mean	0.725	0.833	6.680	10.567
+1 standard deviation	0.725	0.804	6.680	9.278

Notes: Table reports IV estimates of variants of equation (10) in columns 1 and 2 and OLS estimates of a variant of equation (12) in columns 3 and 4. The unit of anlaysis is an economic sub-region (ESR)-year, and the regressions include ESR fixed effects and year fixed effects. All dependent variables are in logs. In all columns hospital expenditures and income are divided by a hospital-utilization weighted measure of population (HUWP) before taking logs. The sample is all Southern states between 1970 and 1990. Note that the results in columns 1 and 3 differ slightly from baseline results in Table 4 because the sample does not include Washington, DC (DC is dropped because there is no data for DC in the 1970s). Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets.

Table 10: Heterogeneity Across Geography and Time

	Panel A: IV Results				
	Dependent Variable: Hospital Expenditures				
	(1)	(2)	(3)	(4)	(5)
log(Income)$_{jt}$	0.723	0.700	0.804	*N/A*	0.853
	(0.214)	(0.368)	(0.633)		(0.439)
	[0.004]	[0.106]	[0.210]		[0.071]
R^2	0.968	0.967	0.956		0.970
N	2065	1070	4915		3547
	Panel B: First Stage Results				
	Dependent Variable: Income				
	(1)	(2)	(3)	(4)	(5)
Oil Reserves$_j$ × log(oil price)$_{t-1}$	9.245	6.237	7.094	1.481	7.966
	(2.271)	(1.655)	(2.375)	(1.882)	(1.930)
	[0.001]	[0.009]	[0.004]	[0.443]	[0.001]
R^2	0.983	0.985	0.982	0.984	0.986
N	2065	1070	4915	3547	3547
F-statistic	16.58	14.21	8.92	0.62	17.04
	Specification				
	(1)	(2)	(3)	(4)	(5)
Years	1970-1990	1970-1990	1970-1990	1970-2005	1970-2005
Geographic sample	South	Southern States w/ Large Oil Wells	All US	South	South
State-specific time trends	N	N	N	N	Y

Notes: Table reports estimates of variants of estimating equation (10) by IV in Panel A and equation (11) by OLS in Panel B. All dependent variables and income are in logs and divided by a hospital-utilization weighted measure of population (HUWP) before taking logs. Unit of analysis is an economic sub-region (ESR)-year in all columns, and all columns include ESR fixed effects and year fixed effects. Bottom rows define the specification variants. The baseline sample is all Southern states between 1970 and 1990. Column 1 reproduces baseline results from column 7 in Table 3 and column 3 in Table 4. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets. Because there is no statistically significant first stage in column 4, the IV results are not reported.

Table 11: Augmented Dickey-Fuller Tests

	Dependent Variable: $\log(\text{oil price})_t$ - $\log(\text{oil price})_{t-1}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\log(\text{oil price})_{t-1}$	0.034	0.005	0.014	0.010	-0.090	-0.156	-0.151	-0.175
	(0.054)	(0.057)	(0.060)	(0.063)	(0.089)	(0.093)	(0.101)	(0.107)
	[0.537]	[0.927]	[0.816]	[0.880]	[0.315]	[0.098]	[0.141]	[0.107]
$\log(\text{oil price})_{t-1}$ - $\log(\text{oil price})_{t-2}$		0.249	0.254	0.264		0.318	0.319	0.351
		(0.158)	(0.160)	(0.167)		(0.156)	(0.159)	(0.166)
		[0.120]	[0.119]	[0.121]		[0.046]	[0.050]	[0.041]
$\log(\text{oil price})_{t-2}$ - $\log(\text{oil price})_{t-3}$			-0.121	-0.123			-0.038	-0.034
			(0.166)	(0.170)			(0.167)	(0.169)
			[0.469]	[0.474]			[0.819]	[0.840]
$\log(\text{oil price})_{t-3}$ - $\log(\text{oil price})_{t-4}$				0.047				0.125
				(0.172)				(0.170)
				[0.786]				[0.467]
t					0.111	0.142	0.141	0.157
					(0.064)	(0.065)	(0.070)	(0.075)
					[0.088]	[0.035]	[0.050]	[0.040]
N	55	54	53	52	55	54	53	52
Dickey-Fuller test statistic	0.621	0.092	0.234	0.151	-1.014	-1.686	-1.498	-1.642
Approximate p-value	0.988	0.966	0.974	0.969	0.942	0.757	0.830	0.776

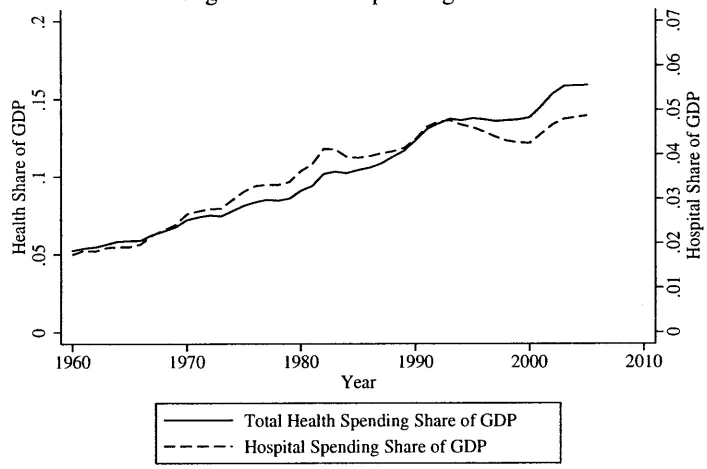
Notes: Table based on annual data on oil prices from 1950 to 2005 (see Figure 2). Standard errors are in parentheses and p-values are in brackets.

Table 12: Short-run versus Long-run Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Income	Hospital Expenditures	Hospital Expenditures	Income	Hospital Expenditures	Hospital Expenditures	Hospital Expenditures
	Baseline	Baseline	Baseline	10-year	10-year	10-year	3-year avg.
	FS	RF		FS	RF		
	OLS	OLS	IV	OLS	OLS	IV	IV
Oil Reserves$_j$ × $\log(\text{oil price})_{t-1}$	9.245	6.680		7.621	6.050		
	(2.271)	(2.099)		(2.643)	(2.628)		
	[0.001]	[0.006]		[0.011]	[0.036]		
$\log(\text{Income})_{jt}$			0.723			0.794	0.826
			(0.214)			(0.411)	(0.231)
			[0.004]			[0.073]	[0.003]
R^2	0.983	0.973	0.968	0.986	0.986	0.981	0.976
N	2065	2065	2065	296	296	296	690
F-statistic	16.577			8.318			

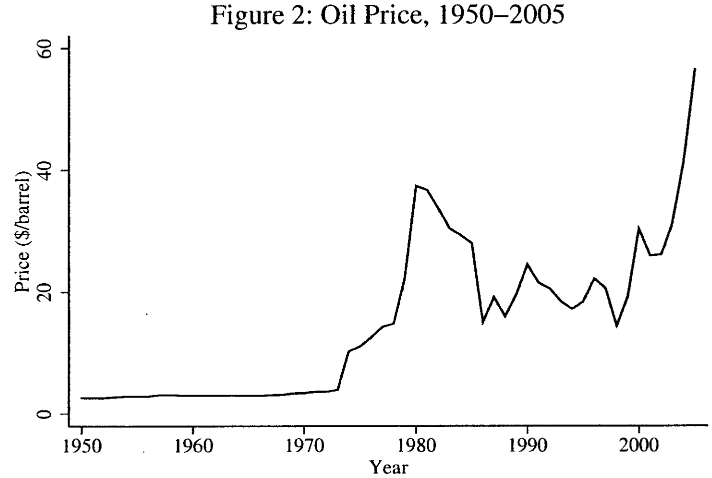
Notes: Table reports results of estimating equations (10), (11) or (12) by OLS or IV as indicated. All dependent variables are in logs. Unit of analysis is an economic sub-region (ESR)-year, and all columns include ESR fixed effects and year fixed effects. In all columns income and hospital expenditures are divided by a hospital-utilization weighted measure of population (HUWP) before taking logs. Columns 1 through 3 are the baseline sample of all Southern states between 1970 and 1990; in columns 4 through 6, only observations from 1970, 1980, and 1990 are included. Column 7 uses 3-year averages of all variables (see Appendix Section B for more details). Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets.

Figure 1: Health Spending Trends



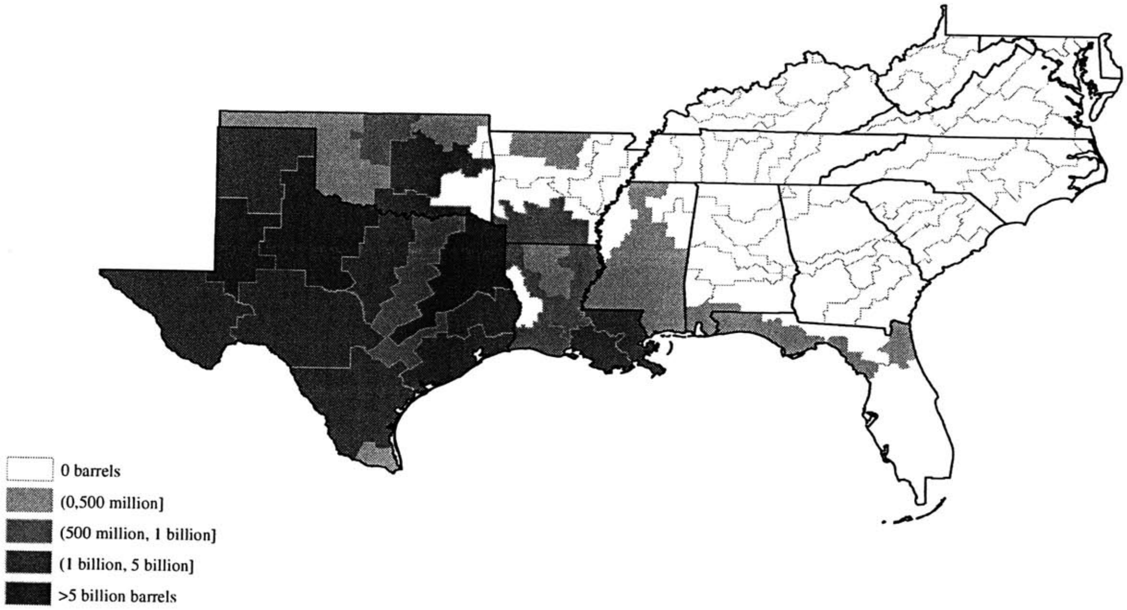
Note: This graph displays the trends in hospital spending from 1960 until 2005. Source: CMS (2006).

Figure 2: Oil Price, 1950–2005



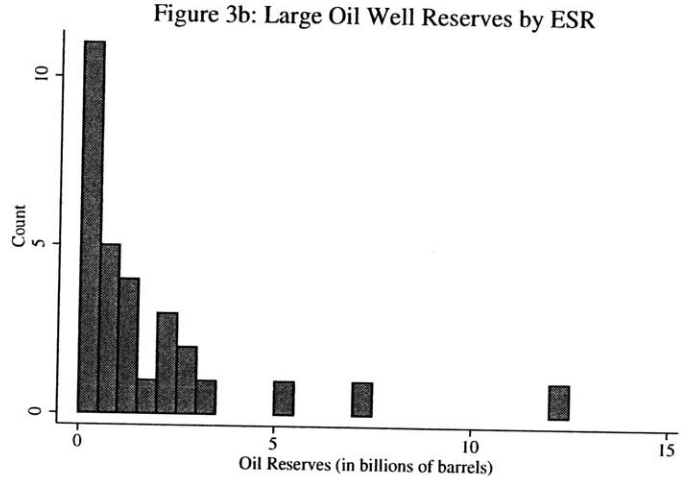
Note: This graph displays the annual average oil price, calculated from the monthly spot prices in the West Texas Intermediate series. The data are available here: `http://research.stlouisfed.org/fred2/series/OILPRICE/downloaddata?cid=98`.

Figure 3a: Map of Large Oil Well Reserves by ESR



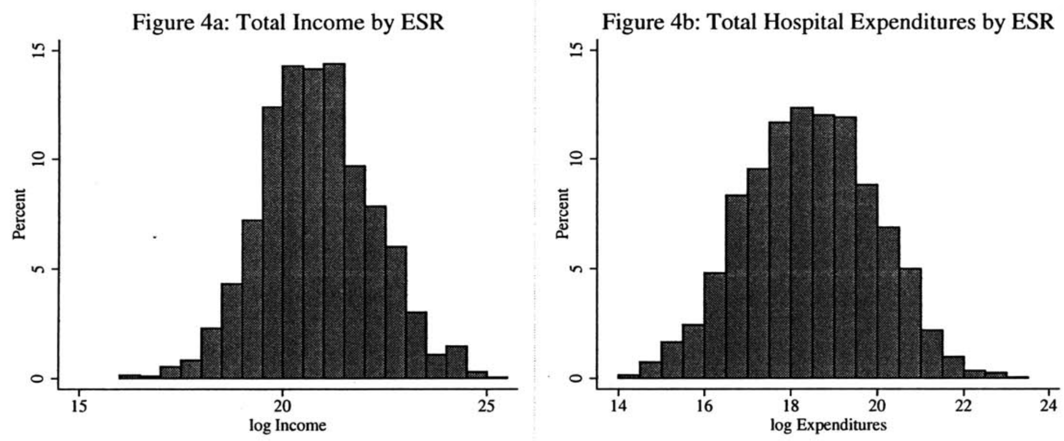
Note: This map displays the total amount of oil in large oil wells for each economic sub-region in the South. Large oil wells are defined as having ever had more than 100 million barrels of oil. The data come from the 2000 Edition of the Oil and Gas Data Book.

Figure 3b: Large Oil Well Reserves by ESR



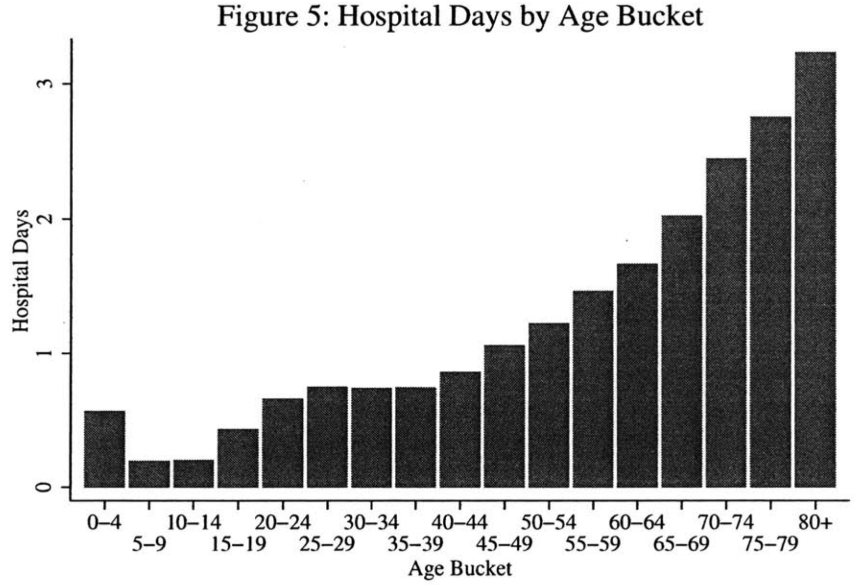
Note: This figure displays the cross-sectional distribution of oil reserves by economic sub-region (ESR) among the ESRs containing large wells. Of the 99 ESRs in the South, 69 ESRs do not have any large oil wells. This figure shows the amount of oil reserves (in billions of barrels) for the 30 ESRs with large oil wells. The data come from the 2000 Edition of the Oil and Gas Data Book.

Figure 4: Aggregate Income and Hospital Expenditure Data



Note: This figure contains histograms of the total income and total hospital expenditures by economic sub-region (ESR). Income is measured using the payroll data from the County Business Patterns (CBP), and the total hospital expenditures come from the American Hospital Association (AHA) Annual Surveys. Both variables are displayed in logs. The data displayed are for ESRs in the South for the years 1970 to 1990.

Figure 5: Hospital Days by Age Bucket



Note: This chart displays the average annual number of hospital days for various age buckets. The data come from the National Health Interview Survey (NHIS) for years 1973 to 1991.

2.5 Appendix A: Induced Innovation Effects

In this Appendix, we present a simple model to illustrate why, given an income elasticity of health expenditure less than one, any induced innovation effects in the health care sector due to rising income are unlikely to be large. We first present a simple model incorporating endogenous technology responses to changes in market size. To economize on space, the reader is referred to Acemoglu (2002, 2007, 2009) or Acemoglu and Linn (2004) for the details (and microfoundations for various assumptions imposed here for simplicity).

Consider an infinite-horizon, continuous-time economy with $g = 1, ..., G$ goods. To communicate the basic ideas, we take expenditures on these goods as given, represented by $[E_g(t)]_{t=0}^{\infty}$ for good g (in terms of some numeraire). We also assume that all of these goods have unit price elasticity (otherwise, we could not take these expenditures as given). We then ask how changes in these expenditure levels affect the types of technologies developed by profit-maximizing firms. These assumptions imply that at time t the demand for good g will be

$$D_g(p_g(t), t) = \frac{E_g(t)}{p_g(t)},$$

where $p_g(t)$. Suppose, in particular, that each good can be supplied in different qualities, denoted by $q_g(t) \in \mathbb{R}_+$, and consumers will purchase whichever variety of the good has the highest price-adjusted quality. That is, among varieties of good g, $g_1,..., g_V$, available in the market, they will choose the one with highest $q_{g_v}(t)/p_{g_v}(t)$. This implies that whichever firm has the highest quality variety for good g at time t will generate revenues equal to $E_g(t)$. Suppose also that all goods, regardless of quality, can be produced at marginal cost equal to 1 (in terms of the numeraire). This implies that the firm with the highest price-adjusted quality for good g at time t (presuming that there is a single such firm) will make profits equal to

$$\pi_g(t) = (p_g(t) - 1)\frac{E_g(t)}{p_g(t)}. \tag{14}$$

Innovation and technological progress are modeled as in the quality ladder models of Aghion and Howitt (1992) and Grossman and Helpman (1991) (see also Acemoglu, 2009, for a textbook treatment). Suppose that starting from leading-edge quality $q_g(t)$ at time t, R&D directed to good g generates (stochastic) innovations for this good. An innovation creates a new leading-edge quality $\lambda q(t)$, where $\lambda > 1$. There is free entry into R&D and each firm has access to an R&D technology that generates a flow rate δ_g of innovation for every dollar spent for research on good g. So if R&D expenditure at time t for good g is $z_g(t)$, the flow rate of innovation is

$$\delta_g z_g(t).$$

Differences in δ_g's introduce the possibility that technological progress is scientifically more difficult for some goods than for others. A firm that makes an innovation has a perpetual patent on the good that it invents, and will be able to sell it until a better good comes to the market.

Consider good g, where current quality is $q_g(t)$. Consumers will purchase from the highest price-adjusted quality and, by definition, the next best firm must have quality $q_g(t)/\lambda$ and can price as low as its marginal cost, 1. This implies that the leading-edge producer must set a limit price

$$p_g(t) = \lambda \text{ for all } g \text{ and } t. \tag{15}$$

Then (14) gives the time t profits of the firm with the leading-edge variety of good g, with quality $q_g(t)$ as

$$\pi_g(q_g(t)) = \frac{\lambda - 1}{\lambda} E_g(t). \tag{16}$$

Firms are forward-looking, and discount future profits at the interest rate r. We assume that this interest rate is constant. The discounted value of profits for firms can be expressed by a standard dynamic programming recursion. $V_g(t \mid q_g)$, the value of a firm that owns the most advanced variety of good g with quality q_g at time t, is

$$rV_g\left(t \mid q_g\right) - \dot{V}_g\left(t \mid q_g\right) = \pi_g\left(q_g\left(t\right)\right) - \delta_g z_g\left(t\right) V_g\left(t \mid q_g\right), \tag{17}$$

where $\pi_g\left(q_g\left(t\right)\right)$ is the flow profits given by (16), and $z_g\left(t\right)$ is R&D effort at time t on this line by other firms. Throughout, we assume that the relevant transversality conditions hold and discounted values are finite. Moreover, because of the standard replacement effect first emphasized by Arrow (1962), the firm with the best technology does not undertake any R&D itself (see, for example, Aghion and Howitt, 1992, Acemoglu, 2009). Intuitively, the value of owning the best technology for good g, $rV_g\left(t \mid q_g\right)$, is equal to the flow profits, $\pi_g\left(q_g\left(t\right)\right)$, plus the potential appreciation of the value, $\dot{V}_g\left(t \mid q_g\right)$, and takes into account that at the flow rate $\delta_g z_g\left(t\right)$ there will be a new innovation, causing the current firm to lose its leading position and to make zero profits thereafter.

Free entry into R&D for developing new technologies for each good implies that there will be entry as long as additional R&D is profitable. Therefore, free entry requires the following complementary slackness condition to hold:

$$\text{if } z_g\left(t\right) > 0, \text{ then } \delta_g V_g\left(t \mid q_g\right) = 1 \text{ for all } g \text{ and } t \tag{18}$$

(and if $z_g\left(t\right) = 0$, $\delta_g V_g\left(t \mid q_g\right) \leq 1$ and there will be no innovation for this good at time t).

An equilibrium in this economy is given by sequences of prices $p_g\left(t\right)\big|_{g=1,..G}$ that satisfy (15), and R&D levels $z_g\left(t\right)\big|_{g=1,..g}$ that satisfy (18) with $V_g\left(\cdot\right)$ given by (17).

An equilibrium is straightforward to characterize. The free entry condition must hold at all t. Supposing that it holds as equality in some interval $[t', t'']$, we can differentiate this equation with respect to time, which yields $\dot{V}_g\left(t \mid q_g\right) = 0$ for all g and t (as long as $z_g\left(t\right) > 0$). Substituting this equation and (18) into (17) yields the levels of R&D effort in the unique equilibrium as

$$z_g(t) = \max\left\{\frac{\delta_g(\lambda - 1)\lambda^{-1}E_g(t) - r}{\delta_g}; 0\right\} \text{ for all } g \text{ and } t. \tag{19}$$

Equation (19) highlights the market size effect in innovation: the greater is expenditures on good g, $E_g(t)$, the more profitable it is to be a supplier of that good, and consequently, there will be greater research effort to acquire this position. In addition, a higher productivity of R&D as captured by δ_g also increases R&D, and a higher interest rate reduces R&D since current R&D expenditures are rewarded by future revenues.

Given equation (19), we can now ask how a rise in overall income in the economy will affect the direction of technological change. Such a change will shift the expenditures from $\left\{[E_g(t)]_{t=0}^{\infty}\right\}_{g=1,\dots,G}$ to $\left\{\left[\tilde{E}_g(t)\right]_{t=0}^{\infty}\right\}_{g=1,\dots,G}$. However, expenditures on some good will increase by more, in particular, those that are "luxury goods" will see their expenditures increase by more. Equation (19) then implies that innovations will be tend to be directed towards those goods.

To highlight the implications of this type of induced technological change for our purposes, suppose that the economy consists of two goods, health care and the "rest". Suppose also that equation (19) leads to positive R&D for both groups of goods. Moreover, let us parameterize expenditures on these two groups of goods as $E_{health}(t) = a_{health}(t)Y(t)$ and $E_{rest}(t) = a_{rest}(t)Y(t)$, where $Y(t)$ is total income (GDP). Our ESR-level estimates imply that, without the induced technology responses, $a_{rest}(t) > a_{health}(t)$, so that with the rising incomes $E_{rest}(t)$ increases more than $E_{health}(t)$. Equation (19) then implies that $z_{rest}(t)$ will increase (proportionately) by more than $z_{health}(t)$, or that $z_{rest}(t)/z_{health}(t)$ will increase. Importantly, this conclusion is independent of the values of the δ_g's as long as they are such that both $z_{rest}(t) > 0$ and $z_{health}(t) > 0$. This result is the basis of our argument that, given the relationship between health care expenditures and income we observe at the ESR level, national-level directed technological change is unlikely to significantly increase the responsiveness of health care expenditures to aggregate income changes.

Equation (19) also highlights the conditions under which this conclusion needs to be modified. If it happens to be the case that $z_{health}(t) = 0$ and $z_{rest}(t) > 0$ to start with, then an increase in $E_{health}(t)$ that is proportionately less than that in $E_{rest}(t)$ may still have a disproportionate effect on innovation in the health care sector by making $z_{health}(t) > 0$. Intuitively, before the changes in expenditures, technological change in the health care sector would have been unprofitable, and as the market size passes a certain threshold (in this case equal to $\delta_g^{-1}(\lambda - 1)^{-1}\lambda r$), innovation jumps up from zero to a positive level. While this is theoretically possible, we believe that it is unlikely to be important in the context of the health care sector, since as discussed earlier in the main text, throughout the 20th century technological change in the health care sector was positive and in fact quite rapid (Cutler and Meara, 2003).

2.6 Appendix B: Robustness

In this Appendix, we provide several robustness checks of our baseline estimates, particularly focusing on whether our causal estimates of the effect of income on health care expenditures might be spurious and whether they may be underestimating the income elasticity of health care expenditures. In the interest of brevity, we focus our discussion on the robustness of our main dependent variable: hospital expenditures. Appendix Table A6 summarizes results from the alternative specifications shown in the previous tables as well as results from the main alternative specifications pursued below, for each of the components of hospital expenditures analyzed in Table 5.

2.6.1 Exclusion Restriction

The exclusion restriction of our IV strategy is that absent oil price changes, ESRs with different levels of oil reserves would have experienced the same proportional changes in hospital expenditures. In Table A3 we explore a variety of alternative specifications designed to investigate the validity of this identifying assumption. As usual, Panel A shows the IV estimates, while Panel B shows the corresponding first-stage results. Column 1 replicates our baseline estimates.

Column 2 shows the results of a natural falsification test: we repeat the baseline analysis of equation (11) (corresponding to column 1), but also include a 5-year *lead* of the instrument, that is, $\log p_{t+5} \times I_j$ (where I_j again denotes oil reserves in ESR j). To the extent that our instrument captures the impact of rising oil prices on the area's income rather than differential trends across areas with different levels of oil reserves, future oil prices should not predict current income changes. Column 2 in Panel B shows that the first-stage relationship is robust to including the lead of the instrument. The coefficient on the lead of the instrument is positive and large (about 60 percent of that on the instrument), though statistically insignificant. The magnitude of this coefficient raises some concerns about potential serial correlation. We explore issues of serial correlation in greater detail in the subsection 2.6.3. To preview, even if there is serial correlation in the first stage, this does not necessarily create a bias in the IV estimates. In addition, our robustness checks in the next subsection show that the statistical and quantitative properties of our estimates are reasonably robust in alternative specifications that explicitly recognize the possibility of serial correlation.

The results from the IV estimates that include the five-year lead of the instrument (both in the first and second stages) are shown in Panel A column 2. The estimate of income elasticity in this specification remains statistically significant and increases somewhat in magnitude relative to the baseline in column 1. The negative (and statistically insignificant) coefficient on the five-year lead of the instrument indicates that our IV estimates are unlikely to be capturing pre-existing trends.

Column 3 shows the results from an alternative check on our identification strategy, in which we additionally control for interactions between oil prices ($\log p_{t-1}$) and fixed ESR characteristics. In particular, we control for separate interactions between log oil prices in year t-1 and each of log hospital expenditures in 1969, log hospital beds in 1969, log population in 1970, log area income in 1970 and log area employment in 1970. This "horse race" between our instrument and other interactions of oil prices and baseline area characteristics is useful for two complementary reasons. First, it provides additional evidence that it is the interaction between oil price shocks and availability of oil reserves leading to the source of income variation that we are exploiting. Second, it indirectly controls for differential pre-existing trends in health expenditures (and income) across ESRs, which are the main threat to our identification strategy. Consistent with the limited differences in various ESR characteristics shown in Table 2, the results of this horse race show that both our first-stage and second-stage estimates are robust in magnitude and precision to the (simultaneous) inclusion of all of these interaction terms. Very similar estimates are obtained when we include each interaction term one by one (not shown).

Column 4 shows the results of adding region-specific linear trends for the three Census regions within the South. Column 5 shows the results of adding state-specific linear trends. These two specifications allow different regions (respectively, different states) within the South to be on different linear time trends. The first stage is reasonably robust. The IV estimates decline considerably in magnitude, and in the case of state specific linear trends, they are no longer statistically significant. Although this last result raises some concerns about the magnitude and precision of our estimates of the income elasticity, if anything, it suggests that our baseline model which does not control for state-specific trends might lead to over-estimates (rather than under-estimates) of this elasticity.

Finally, as another natural and important falsification exercise, we checked the implications of estimating our models on health expenditures data from 1955 through 1969 while assuming that the oil price changes took place 15-years prior (more precisely, the year 1955 is assigned the oil price for 1970, the year 1956 is assigned to the oil price in 1971, and so on through the year 1969 which is assigned to the oil price of 1984).[52] The period before 1970 shows virtually constant oil prices before 1970 (see Figure 2). Therefore, if our identifying assumption is valid, we should not see any differential changes in health expenditures across areas with different oil reserves prior to 1970, and in particular, we should not see more rapid increases in health expenditures in areas with greater oil reserves. Column 6 shows the first-stage and reduced-form results for our baseline specification if we limit it to the 1970 to 1984 period. The first-stage remains as does the reduced form, though the implied IV estimate is about one half the size of our baseline estimate (which uses the entire 1970-1990 period). Column 7 shows the result for the falsification exercise. Reassuringly, this falsification exercise shows no evidence of a significant reduced-form relationship between our instruments and health expenditures; the point estimate is negative (opposite sign from the "actual" estimate in column 6) and not statistically significant. This finding supports the validity of the identifying assumption that, absent changes in oil prices, areas of the South with different levels of oil intensity would have experienced similar trends in their hospital expenditures.

[52]The AHA data do not contain information on hospital expenditures prior to 1955, which is why we could not extend this analysis even further back in time. We report only reduced-form results for this falsification exercise because we do not have income data for the entire period from 1955 to 1969. Our primary source of income data, CBP, extends back annually to 1964 and is available irregularly dating back to 1946. However, before 1970 only first quarter payroll and employment data are available from CBP.

Overall, we read the results in Table A3 as broadly supportive of our identifying assumption.

2.6.2 Alternative Specifications of the Instrument

We also explored the robustness of our results to alternative specifications of the instrument. Table A4 shows the results. Panel A again shows the IV estimates and Panel B shows the corresponding first-stage estimates. Column 1 replicates our baseline first-stage specification, in which the instrument is the interaction of the total oil reserves and the log of the (lagged) oil price, i.e., $\log p_{t-1} \times I_j$, with again I_j measured as oil reserves. The remainder of the columns show results for alternative (plausible) specifications of the instrument; they tend to produce smaller income elasticities than our baseline specification.

Columns 2 and 3 report results using different functional forms for oil prices. Column 2 reports results in which the instrument is constructed as the interaction between the level of (lagged) oil prices and oil reserves (i.e., $p_{t-1} \times I_j$ instead of $\log p_{t-1} \times I_j$ as in our baseline specification). Column 3 reports results when we use the log oil price at time t rather than its one year lag (i.e., $\log p_t \times I_j$ instead of $\log p_{t-1} \times I_j$). With both alternative functional forms for oil prices we continue to estimate strong first stages and statistically significant income elasticities in the second stage that are similar to, though slightly smaller than, our baseline estimate (the income elasticity estimates are 0.49 and 0.64 in columns 2 and 3 respectively, compared to 0.72 in our baseline).

Columns 4 through 6 report results using different ways of measuring the oil intensity of the area. Recall that in our baseline specification we proxied oil intensity of area j by its total (cumulative) oil reserves. Figure 3b shows that the oil reserve distribution is highly skewed and one may be concerned that using the level of oil reserves might give disproportionate weight to the ESRs with the highest oil reserves. Moreover, the effect of oil reserves on the demand for labor, and thus on income, may be nonlinear, with large and very large oil reserves leading to similar effects on income when oil prices rise. Motivated by these considerations, in column 4 we report results with an alternative measure of I_j, where oil reserves are censored at the 95th percentile of oil reserve distribution (the instrument is then constructed by interacting this measure with $\log p_{t-1}$). The results are very similar to the baseline. We continue to estimate a strong first stage, and a statistically significant income elasticity; the estimated income elasticity of 0.632 (standard error = 0.205) is only slightly smaller than the baseline estimate. We also obtain similar estimates if instead we censor oil reserves at the 90th or the 99th percentiles (not shown).

As another check on possible nonlinearities, column 5 measures oil intensity by an indicator variable for whether there are any large oil wells in the ESR (i.e., the instrument is now $\mathbf{1}(I_j > 0)$). The first stage is now slightly weaker (F-statistic of about 8), and the estimated income elasticity rises to 1.10 (standard error = 0.67), but is no longer statistically significantly at the 5 percent level.

Finally, in column 6 we measure oil intensity as the (de-meaned) mining share of employment in the ESR in 1970, interacted with an indicator variable for whether there are any large oil wells in the ESR.[53] Our first stage is now marginally stronger than in the preceding specification (F-statistic of about 11), and we estimate a statistically insignificant income elasticity of 0.860 (standard error = 0.870).

2.6.3 Serial Correlation and Standard Errors

In our baseline model we cluster our standard errors at the state level; the standard errors are therefore computed from a variance-covariance matrix that allows both for arbitrary correlation in residuals across ESRs within a state and for serial correlation at the state or ESR level. However, because we only have 16 states in our baseline (South only) sample, these standard errors may be downward biased due to the relatively small number of clusters (Cameron, Gelbach and Miller, 2008). As a simple robustness check, we computed the standard errors allowing for an arbitrary variance-covariance matrix at the ESR level (rather than the state level). A possible

[53]We include the indicator variable for whether there are any large oil wells because mining employment is defined in the data to include all workers in oil mining, natural gas and coal mining. The indicator for oil wells is included to separate out high mining share non-oil areas (such as coal mining areas of West Virginia).

disadvantage of these standard errors is that they do not allow for correlation across ESRs within the same state, which may be important in practice.[54] Clustering at the ESR level increases the standard errors substantially, so that the first-stage F-statistic is now 5.50 (instead of 16.58 with clustering at the state level). The standard errors for the second stage are also larger, but our IV estimate is still statistically significant at the 6 percent level (results available upon request).

Another strategy to correct for potential biases in the standard errors resulting from the small number of clusters at the state level is the wild bootstrap procedure suggested by Cameron, Gelbach and Miller (2008).[55] We performed wild bootstraps resampling states with replacement. In this case, we find reassuringly similar (indeed somewhat smaller) p-values to our baseline specification with state-level clustering.[56] In particular, using wild bootstraps we find that both the first stage and the second-stage estimates are statistically significant at the less than 1 percent level (results available upon request).

An alternative strategy to address concerns about potential serial correlation is to directly model the dynamics of the error term in our structural equation (10) and then estimate this extended model using instrumental-variables Generalized Least Squares (IV-GLS). In all of our IV-GLS specifications we allow for heteroscedasticity in the second-stage error term; we also experiment with various assumptions regarding the nature of any autocorrelation. The details of the implementation of IV-GLS and the procedure for the computation of the standard errors are discussed in Section C. Table A5 reports the results. Column 1 shows estimates from our baseline specification, but using a subsample of our original data; we limit the sample to the 96 (out of 99) ESRs that have data in the full 21 years from 1970 to 1990. Column 1 verifies that

[54]For example, a boom in an oil-rich ESR may attract in-migration from other ESRs within the same state, reducing total payroll income in these ESRs and also potentially affecting health care expenditures through this and other channels. The result would be a negative correlation in ESR-level residuals within a state.

[55]We thank Doug Miller for suggestions and for providing us with a sample code.

[56]In their Monte Carlo study, Cameron et al find it is important to calculate p-values based on t-statistics rather than parameter estimates. We also computed p-values using parameter estimates, and found these to be even lower (thus leading to more precise results) than the results reported here based on t-statistics.

this has no notable effect on our baseline results. Column 2 reports IV-GLS results assuming a common AR(1) autocorrelation coefficient across all ESRs. Column 3 reports results assuming an AR(2) specification of the residuals with common autocorrelation coefficients. In both specifications the point estimate rises relative to the baseline, but is also considerably less precise. Columns 4 and 5 report results assuming state-specific AR(1) and AR(2) errors respectively. Here the point estimates are very similar to the baseline specification both in magnitude and in precision. Overall we interpret these results as supportive of the robustness of the baseline specification.

As a final strategy to control for serial correlation, columns 6 and 7 include a lagged dependent variable on the right-hand side. In column 6, this model is estimated with ordinary least squares and leads to a long-run elasticity of 0.859 (standard error = 0.213), which is slightly higher than our baseline estimate. However, the least squares estimator in column 6 is inconsistent because of the presence of the lagged dependent variable on the right-hand side. Column 7 estimates the same model using the Arellano-Bond GMM dynamic panel estimator. This GMM procedure estimates the same model in first differences using further lags of the dependent variable as instruments. This leads to a considerably smaller long-run elasticity (= 0.142, standard error = 0.080) than in our baseline. Such smaller long-run elasticities make it even less likely that rising incomes over the past half a century could be the primary driver of the increase in the health share of GDP in the United States.[57]

[57] If we estimate our baseline model in first differences (and thus without further lagged dependent variables on the right-hand side), the results are similar to those reported in column 7 from the GMM procedure. In particular, the point estimate is 0.078 (standard error = 0.106). As we discuss in Section C, heterogeneous adjustment dynamics can introduce significant downward bias in first-difference estimates, and we thus put less weight on this estimate.

2.7 Appendix C: Econometric Issues

In this Appendix, we discuss a number of econometric issues related to the correction for serial correlation and dynamics.

2.7.1 Implementation of IV GLS

We now provided details of the implementation of the IV-GLS estimator used in subsection 2.6.3. In particular, we use the following procedure for this estimation. First, we recover estimates of the residuals ($\hat{\varepsilon}_{jt}$) from the baseline IV specification. Then we use these residuals to estimate the autocorrelation coefficients. For example, when we estimate state-specific autocorrelation coefficients, we run the following regression of $\hat{\varepsilon}_{jt}$ on its lag ($\hat{\varepsilon}_{j,t-1}$) for each state to recover an estimate of the state-specific autocorrelation coefficient, $\hat{\rho}_s$:

$$\hat{\varepsilon}_{jt} = \rho_s \hat{\varepsilon}_{j,t-1} + \xi_{jt}$$

These autocorrelation coefficients are used to create adjusted (LHS and RHS) variables as follows:

$$\tilde{x}_{jt} = x_{jt} - \hat{\rho}_s x_{j,t-1}$$
$$\tilde{y}_{jt} = y_{jt} - \hat{\rho}_s y_{j,t-1}$$

Finally, to adjust for ESR-level heteroskedasticity, we run IV again using the adjusted variables above to recover a new set of residuals ($\hat{\varepsilon}'_{jt}$) and then we create a weighting matrix $\hat{\boldsymbol{\Omega}}$ using these residuals:

$$\hat{\boldsymbol{\Omega}} = \mathbf{I}\left(N_T\right) \otimes \mathbf{diag}\left(\frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}'_{1,t}), \frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}'_{1,t}), \ldots, \frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}'_{J,t})\right)$$

where $\mathbf{I}(\cdot)$ creates an identity matrix and $\mathbf{diag}(\cdot)$ creates a diagonal matrix from a vector. Using this weighting matrix, the IV-GLS estimator is given as follows:

$$\hat{\beta}_{IV-GLS} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{Z}(\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{Z}(\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\Omega}^{-1}\mathbf{y}$$

2.7.2 Performance of different estimators with heterogeneous adjustment dynamics

We now describe results from a simple Monte Carlo study to investigate the performance of various estimators under heterogeneous long-run adjustment dynamics. Our Monte Carlo results suggest that heterogeneous adjustment dynamics may lead traditional fixed effects instrumental variables (FE-IV) estimators to underestimate the true long-run effect. We show that using 3-year averages can reduce this bias. Reassuringly, our 3-year average results are similar to our baseline results (see Table 12, column 7). The remainder of this section describes the set of our Monte Carlo study and our results.

We define the following variables for our simulation:

$$\begin{aligned}
z_{jt} &= N(0,1) \\
a_{jt} &= N(0,1) \\
x_{jt} &= N(0,1) + z_{jt} + a_{jt} \\
\delta_j &= N(0,1) \\
\varepsilon_{jt} &= \rho\varepsilon_{j,t-1} + \xi_{jt} \\
y_{jt} &= x_{jt} + a_{jt} + \delta_j + \varepsilon_{jt}
\end{aligned}$$

where j indexes one of the J panels and t indexes on of the T time periods within a panel. $N(0,1)$ represents an i.i.d. standard normal random variable, z_{jt} represents a valid instrumental variable for x_{jt}, a_{jt} is the unobserved variable that induces a correlation between x_{jt} and the error term in the endogenous fixed effects regression of y_{jt} on x_{jt}, and δ_j is an unobserved fixed effect. ε_{jt} is the error term in the model which follows an AR(1) process ($|\rho| < 1$). We also experiment with serveral other ways to construct y_{jt}:

$$y_{jt} = x_{j,t-1} + a_{jt} + \delta_j + \varepsilon_{jt}$$

$$y_{jt} = \begin{cases} x_{jt} + a_{jt} + \delta_j + \varepsilon_{jt} & \text{if } j < J/2 \\ x_{j,t-1} + a_{jt} + \delta_j + \varepsilon_{jt} & \text{if } j \geq J/2 \end{cases}$$

$$y_{jt} = \begin{cases} x_{jt} + a_{jt} + \delta_j + \varepsilon_{jt} & \text{if } j < J/3 \\ x_{j,t-1} + a_{jt} + \delta_j + \varepsilon_{jt} & \text{if } J/3 \leq j < 2J/3 \\ x_{j,t-2} + a_{jt} + \delta_j + \varepsilon_{jt} & \text{if } j \geq 2J/3 \end{cases}$$

We experimented with the following estimators in in our Monte Carlo study:

1. (FE-IV) Fixed effects IV regression of y_{jt} on x_{jt}, instrumenting x_{jt} by z_{jt}

2. (FD-IV) First differences IV regression of $(y_{jt} - y_{j,t-1})$ on $(x_{jt} - x_{j,t-1})$, instrumenting $(x_{jt} - x_{j,t-1})$ by $(z_{jt} - z_{j,t-1})$

3. (FE-IV-LAG) Fixed effects IV regression of y_{jt} on $x_{j,t-1}$, instrumenting $x_{j,t-1}$ by $z_{j,t-1}$

4. (FE-IV-3YR) Fixed effects IV regression of $\tilde{y}_{js}$ on $\tilde{x}_{js}$ instrumenting $\tilde{x}_{js}$ by $\tilde{z}_{js}$ (where $\tilde{v}_{js}$ denotes the three-year averages of v_{jt} and s represents a three-year groups of years)

5. (FD-IV-3YR) First differences IV regression of $(\tilde{y}_{js} - \tilde{y}_{j,s-1})$ on $(\tilde{x}_{js} - \tilde{x}_{j,s-1})$, instrumenting $(\tilde{x}_{js} - \tilde{x}_{j,s-1})$ by $(\tilde{z}_{js} - \tilde{z}_{js})$

Finally, we choose $J = 10$ and $T = 30$, and we experiment with three values of ρ $(0.1, 0.5, 0.9)$.

The results (based on 500 simulations) are given in Appendix Table A7. There are five panels of results corresponding to each of the five estimators mentioned above. The resuls are the mean of the estimates across each of the simulations and the standard deviation of the parameter estimates (in parentheses underneath). The first panel reports the FE-IV results. As would be expected, the standard deviation of the parameter estimates is larger when there are higher amounts of serial correlation. The second panel reports FD-IV results, where (also as expected) the standard deviation of the parameter estimates goes down as there is more serial correlation. The third panel reports FE-IV-LAG results, and the last two columns report the two sets of 3-year average results (FE-IV-3YR and FD-IV-3YR).

Each panel reports results for the same set of four models. The first row is the standard model where all panels adjust instantly. All estimators except FE-IV-LAG perform very well (the average of the parameter estimates is very close to the true value of 1.000). The second row reports results using a model where all panels take one time period to adjust. For this model the FE-IV and FD-IV results perform very poorly, while FE-IV-LAG unsurprisingly performs optimally. Interestingly, FE-IV-3YR still performs reasonably well, though for all degrees of serial correlation the estimates are roughly 2/3 of the true value.

The final two rows (rows 3 and 4) report results when there is heterogeneity in the adjustment dynamics (where a random set of panels responds instantly and another random set of panels does not respond instantly). For all estimators the results are attentuated away from the true coefficient, but the FE-IV-3YR estimator always performs best, even when there is substantial serial correlation.

We conclude two things from this simulation exercise: (1) heterogeneous adjustment dynamics can lead standard estimators (FE-IV and FD-IV) to underestimate the true long-run effect and (2) estimators using 3-year averages appear to be reasonably robust to a moderate amount of heterogeneity in adjustment dynamics.

References

Acemoglu, Daron. 2002. "Directed Technical Change." *Review of Economic Studies* 69(4): 781-809

Acemoglu, Daron. 2007. "Equilibrium Bias of Technology." *Econometrica* 75(5): 1371-1410.

Acemoglu, Daron, *Introduction to Modern Economic Growth*, Princeton University Press, 2009.

Acemoglu, Daron and Joshua Linn. 2004. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry." *Quarterly Journal of Economics*, 119, 1049-1090.

Aghion, Philippe and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323-362.

Arrow, Kenneth J. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies*, 29(1): 155-173.

Cameron, Colin, Jonah Gelbach and Doug Miller. "Boostrap-Based Improvements for Inference with Clustered Errors" forthcoming, *Review of Economics and Statistics*, 90(3): 414-427.

Grossman, Gene M. and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy.* MIT Press, Cambridge, MA.

Appendix Table A1: Hospital Technologies

Hospital Technology	First Year	Last Year	Years of Data	Fraction Adopted
Emergency Department	1970	1990	21	0.998
Histopathology Services	1970	1990	21	0.964
Home care Program / Department	1970	1990	21	0.701
Hospital Auxiliary	1970	1990	21	0.993
Inhalation Therapy Department (Respiratory)	1970	1990	21	0.993
Occupational Therapy	1970	1990	21	0.852
Physical Therapy Department	1970	1990	21	0.993
Psychiatric Partial Hospitalization Program	1970	1990	21	0.727
X-Ray Therapy	1970	1990	21	0.873
Blood Bank	1970	1990	20	0.993
Open Heart Surgery Facilities	1970	1990	20	0.528
Psychiatric Emergency Services (Outpatient)	1970	1990	20	0.788
Psychiatric Emergency Services	1970	1990	19	0.887
Rehabilitation Outpatient Unit	1970	1990	19	0.764
Organized Outpatient Department	1970	1988	18	0.940
Social Work Department	1970	1989	17	0.966
Cardiac Intensive Care	1970	1985	16	0.970
Family Planning Service	1970	1985	16	0.630
Psychiatric Foster And/Or Home Care	1970	1986	16	0.393
Self Care Unit	1970	1985	16	0.503
Premature Nursery	1970	1985	15	0.943
Rehabilitation Inpatient Unit	1970	1985	15	0.592
Postoperative Recovery Room	1970	1982	13	0.993
Electroencephalography	1970	1981	12	0.921
Hemodialysis / Renal Dialysis (Impatient)	1970	1981	12	0.682
Hemodialysis / Renal Dialysis (Outpatient)	1970	1981	12	0.675
Organ Bank	1970	1981	12	0.337
Pharmacy with FT Registered Pharmacist	1970	1981	12	0.974
Pharmacy with PT Registered Pharmacist	1970	1981	12	0.942
Psychiatric Inpatient Unit	1970	1980	11	0.750
Intensive Care Unit (Mixed)	1970	1979	10	0.973
Cobalt and Radium Therapy	1970	1978	9	0.669
Radium Therapy	1970	1978	9	0.837
Cobalt Therapy	1970	1977	8	0.693
Extended Care Unit	1970	1974	5	0.810
Basic Emergency Department	1970	1970	1	0.975
Major Emergency Department	1970	1970	1	0.743
Provisional Emergency Unit	1970	1970	1	0.962
Radioisoptope Facility	1970	1970	1	0.852
Genetic Counseling Service	1971	1990	20	0.441
Radioisoptope Facility (Diagnostic)	1971	1990	20	0.967
Radioisoptope Facility (Therapeutic)	1971	1990	20	0.836
Volunteer Services Department	1971	1990	20	0.956
Psychiatric Consultation and Education	1971	1986	16	0.799
Burn Care	1971	1985	15	0.472
Speech Therapist Services / Pathology	1972	1990	19	0.877
Clinical Psychologist Services	1972	1986	15	0.847
Dental Services	1972	1985	14	0.968

Podiatrist Services	1972	1985	13	0.796
Chronic Obstructive Pulmonary Disease	1975	1990	16	0.783
Alcohol / Chemical Dependency (Outpatient)	1975	1990	15	0.742
Skilled Nursing or Long Term Care Unit	1975	1985	11	0.852
Alcohol / Chemical Dependency (Impatient)	1975	1985	10	0.723
Neonatal Intensive Care	1976	1985	10	0.743
Pediatric Unit (Impatient)	1977	1978	2	0.951
Patient Representative Services	1978	1990	13	0.958
Abortion Service (Impatient)	1978	1981	4	0.794
Abortion Service (Outpatient)	1978	1981	3	0.638
Radioactive Implants	1979	1990	12	0.811
Megavoltage Radiation Therapy	1979	1990	11	0.781
Computerized Tomography Scanner (Head or Body)	1979	1990	10	0.859
Pediatric Intensive Care	1979	1985	7	0.773
Cardiac Catheterization	1980	1990	11	0.722
Hospice	1980	1990	11	0.715
Recreational Therapy	1980	1990	11	0.869
Ultrasound Facility (Diagnostic)	1980	1990	11	0.976
Kidney Transplant	1980	1990	7	0.327
Organ Transplant (Other than Kidney)	1980	1990	7	0.377
Chaplaincy Services	1980	1985	6	0.987
Electrocardiography	1980	1985	6	1.000
Intermediate Care for Mentally Retarded	1980	1985	6	0.439
Intravenous Admixture Services	1980	1985	6	0.993
Medical/Surgical Acute Care	1980	1985	6	1.000
Medical/Surgical Intensive Care	1980	1985	6	0.998
Newborn Nursery	1980	1985	6	1.000
Obstetrical Care	1980	1985	6	1.000
Other Long-Term Care / Intermediate Care Facility	1980	1985	6	0.838
Pediatric Acute Care	1980	1985	6	1.000
Pharmacy Unit Dose System	1980	1985	6	0.990
Psychiatric Acute Care	1980	1985	6	0.953
Psychiatric Long Term Care	1980	1985	6	0.568
General Surgical Services	1980	1985	5	1.000
General Laboratory Services	1980	1985	4	1.000
Health Science Library	1980	1990	3	0.968
Psychiatric Intensive Care	1980	1982	3	0.679
Ambulance Services	1980	1981	2	0.930
Anesthesia Service	1980	1981	2	1.000
Autopsy Services	1980	1981	2	0.989
C.T. Scanner (Body Unit)	1980	1981	2	0.761
C.T. Scanner (Head Unit)	1980	1981	2	0.570
Cancer/Tumor	1980	1981	2	0.894
Electromyography	1980	1981	2	0.826
Hemodialysis (Home Care/ Mobile Unit)	1980	1981	2	0.464
NeuroSurgery	1980	1981	2	0.769
Physical Rehabilitation	1980	1982	2	0.856
Pulmonary Function Laboratory	1980	1981	2	0.987
Toxicology	1980	1981	2	0.983
Intravenous Therapy	1980	1980	1	0.886
Medical/Surgical Acute Care (Inpatient)	1980	1980	1	0.335
Rehabilitation	1980	1980	1	0.953

Residential Care	1980	1980	1	0.547
Residential Care (Inpatient)	1980	1980	1	0.280
Day Hospital	1981	1987	7	0.822
Pediatric Psychiatric Services	1981	1986	6	0.777
Health Promotion	1981	1985	5	0.964
Optometric Services	1981	1985	5	0.857
Other Special Care	1981	1985	5	0.877
Sheltered Care	1981	1985	5	0.419
Ambulator Surgical Services	1981	1981	1	1.000
Podiatrist Services (Inpatient)	1981	1981	1	0.873
Podiatrist Services (Outpatient)	1981	1981	1	0.835
Hemodialysis Services	1982	1990	9	0.850
Outpatient Surgery	1982	1990	8	1.000
Abortion Services	1982	1985	4	0.825
Pharmacy Services	1982	1985	4	1.000
Comprehensive Geriatric Assessment Services	1983	1990	8	0.805
Nuclear MRI Facility	1983	1990	8	0.542
Psychiatric Liaison Services	1983	1990	8	0.819
Trauma Center	1984	1990	7	0.751
Alcohol / Chemical Acute Care (Inpatient)	1984	1984	1	0.903
Alcohol / Chemical Subacute Care (Inpatient)	1984	1984	1	0.852
Birthing Room	1985	1990	6	0.970
Extracorporeal Shock-Wave Lithotripter	1985	1990	6	0.395
X-Ray (Diagnostic)	1985	1989	5	0.999
Unknown Technology	1985	1985	1	0.678
Adult Day Care	1986	1990	5	0.567
Community Health Promotion	1986	1990	5	0.984
Fertility Counseling	1986	1990	5	0.608
Fitness Center	1986	1990	5	0.746
Geriatric Acute-Care Unit	1986	1990	5	0.754
Occupational Health Services	1986	1990	5	0.869
Patient Education	1986	1990	5	0.992
Respite Care	1986	1990	5	0.803
Sports Medicine Clinic / Service	1986	1990	5	0.775
Sterilization	1986	1990	5	0.945
Women's Center	1986	1990	5	0.762
Worksite Health Promotion	1986	1990	5	0.959
Organ Transplant (Including Kidney)	1986	1989	4	0.467
AIDS Services	1986	1987	2	0.926
Continuing Care Case Management	1986	1987	2	0.773
Contraceptive Care	1986	1987	2	0.646
Genetic Counseling Screening	1986	1987	2	0.532
Satellite Geriatric Clinics	1986	1987	2	0.278
Child Adolescent Psychiatric Services	1987	1990	4	0.872
Geriatric Psychiatric Services	1987	1990	4	0.839
Psychiatric Education	1987	1990	4	0.887
AIDS (Outpatient)	1988	1990	3	0.414
AIDS General Inpatient Care	1988	1990	3	0.980
AIDS/ARC Unit	1988	1990	3	0.247
AIDS/HIV Testing	1988	1990	3	0.969
Alzheimer's Diagnostic Assessment Services	1988	1990	3	0.596
Emergency Response for Elderly	1988	1990	3	0.932

Geriatic Clinic	1988	1990	3	0.496
In Vitro Fertilization	1988	1990	3	0.379
Medicare Certified Distinct Part Skilled Nursing Unit	1988	1990	3	0.886
Organized Social Work Services	1988	1990	3	0.989
Other Skilled Nursing Care	1988	1990	3	0.891
Senior Membership Program	1988	1990	3	0.737
Angioplasty	1989	1990	2	0.708
Arthritis Treatment Center	1989	1990	2	0.485
Emergency Social Work Services	1989	1990	2	0.911
Freestanding Outpatient Center	1989	1990	2	0.686
Hospital Based Outpatient Care Center	1989	1990	2	0.998
Orthopedic Surgery	1989	1990	2	0.972
Outpatient Social Work Services	1989	1990	2	0.939
Bone Marrow Transplant	1990	1990	1	0.301
Cardiac Rehabilitation	1990	1990	1	0.924
Non-Invasive Cardiac Assessment	1990	1990	1	0.970
Positron Emission Tomography Scanner	1990	1990	1	0.267
Single Photo Emission Computed Tomography	1990	1990	1	0.754
Stereotactic Radiosurgery	1990	1990	1	0.415
Tissue Transplant	1990	1990	1	0.432

Notes: This table lists the 172 unique technologies from the AHA annual surveys between 1970 and 1990. For each technology, this table reports the first year the technology appears, the last year the technology appears, and the fraction of economic sub-region (ESR)-year observations that contain at least one hospital that has adopted the technology.

Appendix Table A2: Results Leaving Out Each State in Census South

Panel A: IV Results

Dependent Variable: Total Hospital Expenditures

	All South	Drop AL	Drop AR	Drop DE	Drop FL	Drop FL	Drop GA	Drop KY	Drop LA	Drop MD	Drop MS	Drop NC	Drop OK	Drop SC	Drop TN	Drop TX	Drop WV
$\log(\text{Income})_{jt}$	0.723	0.702	0.725	0.695	0.725	0.694	0.838	0.714	0.706	0.655	0.782	0.677	0.823	0.764	0.680	0.461	0.750
	(0.214)	(0.226)	(0.216)	(0.216)	(0.216)	(0.219)	(0.183)	(0.212)	(0.272)	(0.215)	(0.214)	(0.235)	(0.184)	(0.223)	(0.222)	(0.695)	(0.248)
	[0.004]	[0.008]	[0.005]	[0.006]	[0.005]	[0.007]	[0.000]	[0.005]	[0.021]	[0.009]	[0.003]	[0.012]	[0.001]	[0.004]	[0.009]	[0.518]	[0.009]
R^2	0.968	0.969	0.967	0.969	0.967	0.968	0.967	0.968	0.969	0.970	0.967	0.968	0.967	0.969	0.968	0.974	0.969
N	2065	1877	1918	2044	2054	2002	1900	1897	1939	1981	1939	1918	1897	1939	1918	1813	1939

Panel B: First Stage Results

Dependent Variable: Income

	All South	Drop AL	Drop AR	Drop DE	Drop DC	Drop FL	Drop GA	Drop KY	Drop LA	Drop MD	Drop MS	Drop NC	Drop OK	Drop SC	Drop TN	Drop TX	Drop WV
Oil Reserves$_j$ ×	9.245	9.312	9.386	9.182	9.205	9.236	9.347	9.874	8.997	9.007	9.349	8.660	8.892	9.051	9.164	21.641	8.236
$\log(\text{oil price})_{t-1}$	(2.216)	(2.375)	(2.303)	(2.222)	(2.215)	(2.285)	(2.356)	(2.363)	(2.015)	(2.221)	(2.303)	(2.127)	(1.678)	(2.265)	(2.311)	(4.879)	(1.831)
	[0.001]	[0.002]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.000]	[0.001]	[0.001]	[0.001]	[0.001]
R^2	0.983	0.983	0.983	0.983	0.983	0.983	0.984	0.983	0.984	0.984	0.983	0.983	0.983	0.983	0.982	0.984	0.984
N	2065	1877	1918	2044	2054	2002	1900	1897	1939	1981	1939	1918	1897	1939	1918	1813	1939
F-statistic	17.41	15.37	16.61	17.08	17.26	16.33	15.73	17.45	19.93	16.45	16.48	16.58	28.09	15.97	15.73	19.67	20.24

Notes: Table reports estimates of variants of estimating equation (10) by IV in Panel A and equation (11) by OLS in Panel B. In all specifications income and hospital expenditures are divided by hospital-utilization weighted measure of population (HUWP) and then logged. First column shows results from our baseline sample of all Southern states between 1970 and 1990 (see column 7 of Table 3 and column 3 of Table 4). Subsequent columns show the results when the state specified in the column heading is omitted from the analysis. Unit of observation is an economic sub-region (ESR)-year; all regressions include ESR and year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets.

Appendix Table A3: Examination of Identifying Assumption

	(1) Baseline	(2) 5-year Lead	(3) Horse Race	(4) Region Trends	(5) State Trends	(6) 1970-1984 Subsample	(7) Falsification Test
	Panel A: IV and Reduced Form OLS Results						
	Dependent Variable: Hospital Expenditures						
	IV	IV	IV	IV	IV	RF	RF
$\log(\text{Income})_{jt}$	0.723	0.992	0.697	0.352	0.131		
	(0.214)	(0.306)	(0.283)	(0.192)	(0.118)		
	[0.004]	[0.005]	[0.027]	[0.088]	[0.286]		
Oil Reserves$_j$ ×						4.980	-3.107
log(oil price)$_{t-1}$						(1.656)	(4.044)
						[0.009]	[0.455]
Oil Reserves$_j$ ×		-11.322					
log(oil price)$_{t+5}$		(7.830)					
		[0.169]					
R^2	0.968	0.964	0.970	0.972	0.976	0.966	0.980
N	2065	2065	2054	2065	2065	1471	1487
	Panel B: First Stage Results						
	Dependent Variable: Income						
Oil Reserves$_j$ ×	9.245	8.186	8.219	11.722	13.774	14.172	
log(oil price)$_{t-1}$	(2.271)	(2.157)	(2.387)	(3.004)	(3.951)	(3.481)	
	[0.001]	[0.002]	[0.004]	[0.001]	[0.003]	[0.001]	
Oil Reserves$_j$ ×		4.821					
log(oil price)$_{t+5}$		(3.291)					
		[0.164]					
R^2	0.983	0.983	0.984	0.984	0.985	0.986	
N	2065	2065	2054	2065	2065	1471	
F-statistic	16.577	14.396	11.853	15.222	12.154	16.571	

Notes: Table reports results from estimating variants of equation (10) by IV in Panel A, except in columns 6 and 7 which show variants of equation (12) estimated by OLS in Panel A; table reports results from estimating variants of equation (11) by OLS in Panel B. All dependent variables are in logs. In all columns hospital expenditures and income are divided by a hospital-utilization weighted measure of population (HUWP) before taking logs. Unit of observation is an economic sub-region (ESR)-year, and all columns include ESR and year fixed effects. In columns 1 through 5 the sample is all Southern states between 1970 and 1990. Column 1 reproduces baseline results (see column 7 of Table 3 and column 3 of Table 4). Column 2 includes a 5-year lead of the instrument as a control variable. Column 3 includes several additional interaction terms as control variables in a "horse race"; the interaction terms are the log oil price interacted with each of the following variables: hospital expenditures in 1969, hospital beds in 1969, population in 1970, wage bill in 1970, and employment in 1970. Column 4 adds region-specific linear time trends for the three Census regions in the South. Column 5 includes state-specific linear time trends for the 16 Southern states. Column 6 produces the first stage and reduced form results for 1970 to 1984 as comparison to the falsification test in column 7, which "grafts" the same oil price series in 1970 to 1984 onto the hospital data in 1955 to 1969. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets.

Appendix Table A4: Alternative Specifications of Instrument

Panel A: IV Results

Dependent Variable: Hospital Expenditures

	(1)	(2)	(3)	(4)	(5)	(6)
$\log(\text{Income})_{jt}$	0.723	0.491	0.640	0.632	1.095	0.860
	(0.214)	(0.145)	(0.194)	(0.205)	(0.670)	(0.870)
	[0.004]	[0.004]	[0.005]	[0.008]	[0.123]	[0.339]
R^2	0.968	0.971	0.969	0.970	0.962	0.966
N	2065	2065	2065	2065	2065	2065

Panel B: First Stage Results

Dependent Variable: Income

	(1)	(2)	(3)	(4)	(5)	(6)
Oil Reserves$_j$ ×	9.245					
$\log(\text{oil price})_{t-1}$	(2.216)					
	[0.001]					
Oil Reserves$_j$ ×		0.886				
oil price$_{t-1}$		(0.200)				
		[0.000]				
Oil Reserves$_j$ ×			10.080			
$\log(\text{oil price})_t$			(2.467)			
			[0.001]			
max(Oil Reserves,				12.646		
95th percentile) ×				(2.523)		
$\log(\text{oil price})_{t-1}$				[0.000]		
1{Oil Reserves > 0} ×					0.041	
$\log(\text{oil price})_{t-1}$					(0.014)	
					[0.012]	
1{Oil Reserves > 0} ×						0.808
Mining share of labor force in 1970 ×						(0.240)
$\log(\text{oil price})_{t-1}$						[0.004]
R^2	0.983	0.984	0.983	0.983	0.984	0.983
N	2065	2065	2065	2065	2065	2065
F-statistic	17.41	19.71	16.69	25.12	8.22	11.36

Notes: Table reports estimates of variants of estimating equation (10) by IV in Panel A and equation (11) by OLS in Panel B. The specifications vary in their definition of the instrument, which is given in the left-hand column of Panel B. Unit of analysis is an economic sub-region (ESR)-year. All dependent variables are in logs. In all columns hospital expenditures and income are divided by a hospital-utilization weighted measure of population (HUWP) before taking logs. The sample is ESRs in Southern states between 1970 and 1990. Column 1 reproduces baseline results (see column 7 of Table 3 and column 3 of Table 4). 1(Oil Reserves > 0) is an indicator variable for whether the ESR has any large oil wells. All columns include ESR fixed effects and year fixed effects. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets.

Appendix Table A5: IV-GLS and Lagged Depedendent Variable

	Dependent Variable: Hospital Expenditures						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline IV	IV-GLS	IV-GLS	IV-GLS	IV-GLS	Lagged Dep. Var. IV	Arellano-Bond IV
Within-panel serial correlation	Cluster at State	Common AR(1)	Common AR(2)	State-specific AR(1)	State-specific AR(2)	Cluster at State	Cluster at State
$\log(\text{Income})_{jt}$	0.697	0.963	1.111	0.724	0.770	0.491	0.120
(A)	(0.216)	(0.505)	(0.681)	(0.263)	(0.287)	(0.135)	(0.067)
	[0.006]	[0.057]	[0.103]	[0.006]	[0.007]	[0.002]	[0.075]
log(Total Hospital Exp.)$_{t-1}$						0.426	0.154
(B)						(0.088)	(0.047)
						[0.000]	[0.001]
Implied long-run effect						0.856	0.142
$(A/(1-B))$						(0.214)	(0.080)
						[0.001]	[0.077]
N	2016	2016	2016	2016	2016	1966	1966

Notes: Table reports results from estimating variants of equation (10) by IV. The sample is all Southern states between 1970 and 1990. Unit of observation is an economic sub-region (ESR)-year. All specifications include ESR fixed effects and year fixed effects. In all columns, income and hospital expenditures are divided by a hospital-utilization weighted measure of population (HUWP) before taking logs. For columns 1 through 5, the baseline sample is modified to only include the 96 (of 99) ESRs with data for all 21 years between 1970 and 1990. Column 1 produces baseline IV results with this modified sample. Columns 2 through 5 report IV-GLS results. In column 2, ρ_1 is estimated to be 0.585. In column 3, ρ_1 is estimated to be 0.508 and ρ_2 is estimated to be 0.127. In column 4, ρ_1 is estimated separately by state; estimated values of ρ_1 range from 0.155 to 0.887 with mean 0.604 and s.d. 0.240. In column 5, ρ_1 and ρ_2 are estimated separately by state; estimated values of ρ_1 range from 0.118 to 0.747 with mean 0.487 and s.d. 0.200, and estimated values of ρ_2 range from 0.041 to .341 with mean 0.192 and s.d. 0.083. Column 6 includes a lagged dependent variable as a control. Column 7 uses the Arellano-Bond GMM dynamic panel estimator. In columns 6 and 7 the standard error on the implied long-run effect is estimated using the delta method.

Appendix Table A6: Replication of Robustness Analysis for Other Dependent Variables

Dependent Variable:	(1) Total Hospital Expenditures	(2) Total Hospital Payroll	(3) FTE	(4) RN/ (RN+LPN)	(5) Admissions	(6) In-Patient Days	(7) Beds	(8) Number of Hospitals	(9) Number of Technologies	(10) Open-Heart Surgery	(11) Radioisotope Therapy
Panel A: Baseline Results (reproduced from Table 5)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.723	0.934	0.039	0.329	-0.430	-1.034	-0.698	-0.552	-0.132	-3.163	1.083
	(0.214)	(0.233)	(0.222)	(0.089)	(0.193)	(0.488)	(0.455)	(0.358)	(0.221)	(11.334)	(2.575)
	[0.004]	[0.001]	[0.862]	[0.002]	[0.042]	[0.051]	[0.146]	[0.144]	[0.558]	[0.169]	[0.545]
N	2065	2064	2065	1576	2065	1967	2065	2065	2065	849	262
Panel B: No Population Adjustment (see Table 4, column (4))											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.801	0.953	0.311	0.266	-0.025	-0.451	-0.217	-0.395	-0.095	-1.154	1.537
	(0.155)	(0.167)	(0.175)	(0.072)	(0.138)	(0.324)	(0.295)	(0.262)	(0.161)	(3.929)	(3.467)
	[0.000]	[0.000]	[0.096]	[0.002]	[0.861]	[0.184]	[0.474]	[0.152]	[0.564]	[0.187]	[0.259]
N	2065	2064	2065	1576	2065	1967	2065	2065	2065	849	262
Panel C: Per Capita Population Adjustment (see Table 4, column (5))											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.665	0.920	-0.161	0.415	-0.728	-1.468	-1.053	-0.667	-0.160	-2.450	1.894
	(0.263)	(0.282)	(0.297)	(0.112)	(0.265)	(0.645)	(0.584)	(0.447)	(0.272)	(9.014)	(3.297)
	[0.023]	[0.005]	[0.595]	[0.002]	[0.015]	[0.038]	[0.092]	[0.156]	[0.565]	[0.194]	[0.278]
N	2065	2064	2065	1576	2065	1967	2065	2065	2065	849	262
Panel D: State-level Results (see Table 4, column (6))											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.550	0.865	-0.119	0.326	-0.333	0.258	0.294	-0.028	0.819	N/A	N/A
	(0.230)	(0.187)	(0.181)	(0.095)	(0.187)	(0.204)	(0.212)	(0.287)	(1.672)		
	[0.030]	[0.000]	[0.520]	[0.004]	[0.095]	[0.226]	[0.186]	[0.924]	[0.631]		
N	326	326	326	251	326	311	326	326	322		
Panel E: State GSP instead of Income, Census South (see Table 7, column (2))											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.451	0.709	-0.098	0.274	-0.273	0.212	0.241	-0.023	0.673	N/A	N/A
	(0.160)	(0.126)	(0.154)	(0.100)	(0.160)	(0.165)	(0.172)	(0.236)	(1.323)		
	[0.013]	[0.000]	[0.534]	[0.015]	[0.109]	[0.218]	[0.182]	[0.924]	[0.619]		
N	326	326	326	251	326	311	326	326	322		
Panel F: State-level Results, All U.S. (see Table 7, column (3))											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.740	0.863	-0.086	0.407	-0.188	-0.009	0.109	0.098	1.054	N/A	N/A
	(0.359)	(0.398)	(0.244)	(0.112)	(0.342)	(0.268)	(0.395)	(0.295)	(1.367)		
	[0.045]	[0.035]	[0.726]	[0.001]	[0.585]	[0.973]	[0.783]	[0.741]	[0.444]		
N	1015	1015	1015	777	1015	967	1015	1015	1011		
Panel G: State GSP instead of Income, All U.S. (see Table 7, column (4))											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.568	0.662	-0.066	0.331	-0.144	-0.007	0.084	0.075	0.812	N/A	N/A
	(0.263)	(0.298)	(0.189)	(0.102)	(0.264)	(0.205)	(0.303)	(0.225)	(1.023)		
	[0.036]	[0.031]	[0.728]	[0.002]	[0.588]	[0.973]	[0.783]	[0.740]	[0.431]		
N	1015	1015	1015	777	1015	967	1015	1015	1011		
Panel H: Drop States with No Large Oil Wells (see Table 10, column (2))											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.700	0.852	-0.042	0.487	-0.630	-2.089	-1.816	-1.098	-0.047	-2.760	2.996
	(0.368)	(0.378)	(0.544)	(0.124)	(0.305)	(0.578)	(0.626)	(0.513)	(0.237)	(22.066)	(10.381)
	[0.106]	[0.065]	[0.940]	[0.008]	[0.085]	[0.011]	[0.027]	[0.076]	[0.850]	[0.299]	[0.377]
N	1070	1070	1070	815	1070	1019	1070	1070	1070	399	128
Panel I: ESR-level Results, All U.S. (see Table 10, column (3))											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.804	0.882	0.030	0.503	-0.330	-0.598	-0.595	-0.383	-0.191	-2.135	1.292
	(0.633)	(0.557)	(0.352)	(0.126)	(0.406)	(0.439)	(0.605)	(0.320)	(0.325)	(9.394)	(7.234)
	[0.210]	[0.120]	[0.932]	[0.000]	[0.420]	[0.180]	[0.330]	[0.237]	[0.560]	[0.302]	[0.168]
N	4915	4914	4915	3749	4915	4681	4915	4915	4915	1906	503
Panel J: 1970-2005 + state-specific linear time trends (see Table 10, column (5))											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.853	0.935	0.591	0.281	-0.038	-0.486	-0.064	-0.650	0.039	N/A	N/A
	(0.439)	(0.565)	(0.761)	(0.063)	(0.197)	(0.230)	(0.350)	(0.410)	(0.110)		
	[0.071]	[0.119]	[0.450]	[0.000]	[0.850]	[0.052]	[0.858]	[0.134]	[0.729]		
N	3547	3546	3547	3058	3547	3449	3547	3547	2164		
Panel K: Decadal Panel (see Table 12, column (6))											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{it}$	0.794	0.921	0.071	0.625	-0.510	-1.562	-0.771	-0.728	-0.311	N/A	N/A
	(0.411)	(0.541)	(0.548)	(0.250)	(0.306)	(0.846)	(0.689)	(0.491)	(0.355)		
	[0.073]	[0.109]	[0.898]	[0.025]	[0.117]	[0.085]	[0.281]	[0.159]	[0.396]		
N	296	296	296	296	296	296	296	296	296		

	Panel L: 3-year Averages (see Table 12, column (7))										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{jt}$	0.826	0.954	0.235	0.259	-0.369	-0.719	-0.614	-0.635	-0.162	N/A	N/A
	(0.231)	(0.216)	(0.169)	(0.101)	(0.175)	(0.429)	(0.434)	(0.371)	(0.255)		
	[0.003]	[0.001]	[0.185]	[0.022]	[0.052]	[0.115]	[0.178]	[0.107]	[0.534]		
N	690	690	690	592	690	690	690	690	690		
	Panel M: Include 5-year Lead (see Table 13, column (2))										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{jt}$	0.992	1.188	-0.023	0.360	-0.563	-0.979	-0.824	-0.761	-0.175	N/A	N/A
	(0.306)	(0.263)	(0.224)	(0.086)	(0.271)	(0.588)	(0.592)	(0.518)	(0.321)		
	[0.005]	[0.000]	[0.918]	[0.001]	[0.056]	[0.117]	[0.185]	[0.163]	[0.594]		
N	2065	2064	2065	1576	2065	1967	2065	2065	2065		
	Panel N: Horse Race (see Table 13, column (3))										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{jt}$	0.697	0.748	0.117	0.284	-0.617	-0.380	0.084	-0.676	0.314	N/A	N/A
	(0.283)	(0.284)	(0.275)	(0.142)	(0.278)	(0.346)	(0.318)	(0.397)	(0.226)		
	[0.027]	[0.020]	[0.678]	[0.065]	[0.044]	[0.290]	[0.794]	[0.110]	[0.187]		
N	2054	2053	2054	1565	2054	1956	2054	2054	2054		
	Panel O: Region-specific linear time trends (see Table 13, column (4))										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{jt}$	0.352	0.439	0.053	0.262	-0.014	0.279	-0.740	-0.382	-0.040	-2.699	2.023
	(0.192)	(0.254)	(0.287)	(0.062)	(0.127)	(0.071)	(0.371)	(0.210)	(0.134)	(11.844)	(6.711)
	[0.088]	[0.104]	[0.855]	[0.001]	[0.913]	[0.001]	[0.065]	[0.090]	[0.769]	[0.312]	[0.239]
N	2065	2064	2065	1576	2065	1967	2065	2065	2065	849	262
	Panel P: Oil price in levels (see Table 14, column (2))										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{jt}$	0.491	0.648	-0.038	0.230	-0.258	-0.657	-0.506	-0.393	-0.110	-1.392	2.089
	(0.145)	(0.135)	(0.127)	(0.061)	(0.119)	(0.298)	(0.294)	(0.243)	(0.159)	(8.824)	(3.380)
	[0.004]	[0.000]	[0.768]	[0.002]	[0.047]	[0.044]	[0.106]	[0.127]	[0.499]	[0.479]	[0.191]
N	2065	2064	2065	1576	2065	1967	2065	2065	2065	849	262
	Panel Q: Oil price at time t instead of t-1 (see Table 14, column (3))										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{jt}$	0.640	0.823	0.103	0.342	-0.315	-0.996	-0.605	-0.469	-0.173	-3.064	2.016
	(0.194)	(0.233)	(0.222)	(0.104)	(0.171)	(0.458)	(0.403)	(0.306)	(0.210)	(13.207)	(7.030)
	[0.005]	[0.003]	[0.649]	[0.005]	[0.085]	[0.046]	[0.154]	[0.147]	[0.423]	[0.240]	[0.242]
N	2065	2064	2065	1576	2065	1967	2065	2065	2065	849	262
	Panel R: max(Oil Reserves, 95th Percentile) (see Table 14, column (4))										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{jt}$	0.632	0.888	-0.032	0.336	-0.410	-1.025	-0.711	-0.459	-0.145	-2.717	2.003
	(0.205)	(0.227)	(0.206)	(0.095)	(0.187)	(0.482)	(0.454)	(0.335)	(0.225)	(10.449)	(3.490)
	[0.008]	[0.001]	[0.879]	[0.003]	[0.045]	[0.050]	[0.138]	[0.190]	[0.527]	[0.237]	[0.255]
N	2065	2064	2065	1576	2065	1967	2065	2065	2065	849	262
	Panel S: Has Large Oil Wells Dummy (see Table 14, column (5))										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{jt}$	1.095	1.377	0.076	0.297	-0.169	0.078	0.014	0.384	-0.322	0.597	-0.669
	(0.670)	(0.661)	(0.507)	(0.210)	(0.416)	(0.753)	(0.681)	(0.505)	(0.343)	(6.238)	(4.904)
	[0.123]	[0.055]	[0.883]	[0.177]	[0.691]	[0.919]	[0.984]	[0.459]	[0.362]	[0.458]	[0.631]
N	2065	2064	2065	1576	2065	1967	2065	2065	2065	849	262
	Panel T: Oil Wells Dummy x Mining Share of Labor Force in 1970 (see Table 14, column (6))										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{jt}$	0.860	0.975	-0.219	0.189	-0.391	-0.236	0.005	-0.113	0.344	-1.141	-0.212
	(0.870)	(0.805)	(0.484)	(0.162)	(0.450)	(0.877)	(0.871)	(0.584)	(0.604)	(9.114)	(3.217)
	[0.339]	[0.245]	[0.658]	[0.262]	[0.399]	[0.792]	[0.995]	[0.849]	[0.577]	[0.278]	[0.869]
N	2065	2064	2065	1576	2065	1967	2065	2065	2065	849	262
	Panel U: Lagged Dependent Variable (see Table 15, column (6))										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
log(Income)$_{jt}$	0.856	0.951	0.157	0.048	-0.373	-0.505	-0.638	-0.611	-0.114	N/A	N/A
	(0.214)	(0.181)	(0.156)	(0.019)	(0.168)	(0.518)	(0.414)	(0.353)	(0.239)		
	[0.001]	[0.000]	[0.331]	[0.024]	[0.042]	[0.344]	[0.144]	[0.104]	[0.641]		
N	1963	1961	1963	988	1963	1768	1963	1963	1963		

Notes: This table shows robustness results across all of the dependent variables in Table 5 (Panel A reproduces baseline results in Table 5 for comparison). The table and column number in each panel heading references the specification that is being shown; see notes in main tables for details on the various specifications. In all panels, the first column replicates the robustenss analysis shown in the referenced table for total hospital expenditures. In all panels, the dependent variable in columns 4, 8, and 9 is not adjusted for population. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each state over time, are in parentheses and p-values are in brackets. In some panels, there is not enough variation to estimate the Cox proportional hazard models in columns (10) and (11); we place "N/A" in these cells. We do not include robustness tests for the IV-GLS results in Table 15 because several of the alternative dependent variables are missing data for various years, making estimation of the AR(1) and AR(2) coefficients difficult because of the "gaps" in the panel data set. The results reported in Panel U (lagged dependent variable specification) are the implied long-run effects.

Appendix Table A7: Monte Carlo Simulation Results

	FE-IV			FD-IV			FE-IV-LAG			FE-IV-3YR			FD-IV-3YR		
	ρ=0.1	ρ=0.3	ρ=0.9	ρ=0.1	ρ=0.3	ρ=0.9	ρ=0.1	ρ=0.3	ρ=0.9	ρ=0.1	ρ=0.3	ρ=0.9	ρ=0.1	ρ=0.3	ρ=0.9
$y_{jt} = x_{jt} + a_{jt} + \delta_j + \varepsilon_{jt}$	1.008	1.009	1.012	1.010	1.009	1.009	-0.041	-0.038	-0.033	1.020	1.026	1.032	1.029	1.032	1.034
	(0.094)	(0.097)	(0.127)	(0.111)	(0.102)	(0.095)	(0.111)	(0.117)	(0.141)	(0.181)	(0.220)	(0.352)	(0.228)	(0.243)	(0.245)
$y_{jt} = x_{j,t-1} + a_{jt} + \delta_j + \varepsilon_{jt}$	-0.033	-0.031	-0.029	-0.490	-0.491	-0.491	0.993	0.996	1.001	0.626	0.632	0.638	0.496	0.499	0.501
	(0.138)	(0.143)	(0.158)	(0.127)	(0.120)	(0.115)	(0.089)	(0.095)	(0.124)	(0.228)	(0.260)	(0.378)	(0.291)	(0.310)	(0.324)
$y_{jt} = x_{j,t} + a_{jt} + \delta_j + \varepsilon_{jt}$	0.482	0.484	0.486	0.253	0.253	0.253	0.479	0.482	0.486	0.810	0.816	0.821	0.743	0.746	0.748
or $x_{j,t-1} + a_{jt} + \delta_j + \varepsilon_{jt}$	(0.135)	(0.139)	(0.159)	(0.155)	(0.149)	(0.146)	(0.146)	(0.148)	(0.166)	(0.229)	(0.258)	(0.371)	(0.294)	(0.312)	(0.322)
$y_{jt} = x_{jt} + a_{jt} + \delta_j + \varepsilon_{jt}$	0.307	0.309	0.311	0.161	0.161	0.160	0.309	0.311	0.316	0.626	0.632	0.637	0.480	0.483	0.486
or $x_{j,t-1} + a_{jt} + \delta_j + \varepsilon_{jt}$	(0.139)	(0.144)	(0.163)	(0.153)	(0.149)	(0.146)	(0.147)	(0.151)	(0.168)	(0.257)	(0.284)	(0.388)	(0.313)	(0.330)	(0.340)
or $x_{j,t-2} + a_{jt} + \delta_j + \varepsilon_{jt}$															

Notes: This table reports results from the Monte Carlo study described in Appendix (Section C). Each cell displays the mean of the parameter estimates from 500 simulations; standard deviation of parameter estimates is reported below in parentheses.

Chapter 3

Does the Moral Hazard Cost of Unemployment Insurance Vary with the Local Unemployment Rate? Theory and Evidence[1]

(Joint work with Kory Kroft, UC-Berkeley)

It is commonly accepted that higher unemployment benefits prolong unemployment durations (Moffitt 1985, Meyer 1990, Chetty 2008). Most of the evidence for this "moral hazard effect" comes from empirical studies that do not distinguish between changes in benefits when local labor market conditions are good and changes in benefits when local labor market conditions are poor.[1] If the moral hazard cost of Unemployment Insurance (UI) depends on local labor market conditions, this may imply that optimal UI benefits should respond to shifts in local labor demand. However,

[1]We would like to thank Jonathan Guryan for providing excellent comments. Notowidigdo gratefully acknowledges the National Institute of Aging (NIA grant number T32-AG000186) for financial support.

[1]Chetty (2008) shows that it is misleading to interpret the behavioral response to UI benefits as a pure moral hazard effect, as part of the observed response could be coming through liquidity effects. In Section 3.2.1, we investigate the importance of liquidity effects and find no evidence that accounting for liquidity effects significantly alters our main results.

there exists little empirical evidence on measuring how local labor market conditions affect the moral hazard cost of UI, since many of the studies that conduct a welfare analysis of UI do not consider whether and to what extent UI benefits should vary with local labor market conditions (Baily 1978, Chetty 2006, Chetty 2008, Shimer and Werning 2007, Kroft 2008).[2] As Alan Krueger and Bruce Meyer (2002, p64-65) remark:

> [F]or some programs, such as UI, it is quite likely that the adverse incentive effects vary over the business cycle. For example, there is probably less of an efficiency loss from reduced search effort by the unemployed during a recession than during a boom. As a consequence, it may be optimal to expand the generosity of UI during economic downturns ... Unfortunately, this is an area in which little empirical research is currently available to guide policymakers.

Similarly, the Congressional Budget Office writes that the availability of long-term unemployment benefits "could dampen people's efforts to look for work, [but that concern] is less of a factor when employment opportunities are expected to be limited for some time."[3]

In this paper, we conduct both positive and normative economic analyses to investigate how local labor market conditions affect the moral hazard cost of UI. On the positive side, we consider a standard job search model and show that the model implies a steady-state relationship between the disincentive effect of UI and the unemployment rate. We first consider workers who set a reservation wage and face an exogenous arrival rate of job offers. In this version of the model, the relationship between the unemployment rate and elasticity of duration with respect to the UI benefit level is theoretically ambiguous; however, when we calibrate the model using realistic parameter values selected from the literature, the duration elasticity

[2]Nicholson and Needels (2006) discuss how worsening labor market conditions in the U.S. in the 1970s and 1980s triggered large, policy-driven, increases in benefit payments.

[3]The CBO quote is pulled from the following URL: http://www.washingtonpost.com/wp-dyn/content/article/2010/03/08/AR2010030804927_pf.html.

is positively correlated with the unemployment rate.[4] This analysis suggests that the moral hazard cost of UI increases with the unemployment rate, contrary to the speculation of Krueger and Meyer (2002) as well as existing UI policy in the U.S. and many other developed countries.

We extend the search model to encapsulate the more realistic scenario where workers affect the job finding rate by increasing search effort. In this model with an endogenous job offer arrival rate, the elasticity of unemployment duration with respect to the UI benefits is the sum of behavioral responses of (a) reservation wages and (b) search effort. We show that whether moral hazard rises or falls with the unemployment rate depends on the relative importance of these two behavioral channels.

Recent empirical work on the behavioral responses to social insurance programs find that more generous benefits do not lead to higher wages (see Card, Chetty, and Weber 2007). Given that higher UI benefits raise durations, this leads us to suspect that the search effort channel is empirically more important than the reservation wage channel. We examine this question by calibrating the search model with endogenous search effort and considering how variation in local labor market conditions affects the duration elasticity. For different ranges of parameter values, the elasticity can be either positively or negatively related to the unemployment rate. This ambiguity is coming entirely through the search channel – the reservation wage component of the duration elasticity is always increasing with the unemployment rate. We thus conclude from our model and calibrations that the relationship between the duration elasticity and the local unemployment rate is ultimately an empirical question.

To empirically test how the duration elasticity varies with the local unemployment rate, we exploit variation in UI benefit levels within states over time and interact the effect of UI benefit generosity with the state unemployment rate.[5] Our findings

[4] Additionally, we show that we can resolve the theoretical ambiguity by making assumptions on the distribution of wages. If the distribution of wages has a non-increasing hazard rate (as would be the case if wage offers had a Pareto distribution), then the duration elasticity will be increasing in the unemployment rate.

[5] In ongoing work we are constructing variation in state unemployment rates that is driven by plausibly exogenous shifts in local labor demand by following the procedure in Bartik (1991).

indicate that the elasticity of unemployment duration with respect to UI benefits is significantly lower when the local unemployment rate is high. In our preferred specification, the elasticity of unemployment duration with respect to UI benefits is 0.741 (s.e. 0.340) at the mean unemployment rate. However, a one standard deviation increase in the unemployment rate (an increase of 1.68 percentage points) reduces the magnitude of the duration elasticity by 0.239 to 0.502 (a decline in magnitude of 32.3%). To interpret this finding as evidence that the moral hazard cost of UI falls with the unemployment rate, we conduct a variety of robustness tests to address concerns that the interaction effect we estimate is driven by compositional changes, unobserved trends, sample selection, and liquidity effects, and find no evidence that any of these concerns are primarily responsible for our effect. We therefore conclude that the association between the duration elasticity and the local unemployment rate indicates that the moral hazard cost of UI varies systematically with local labor market conditions.

Finally, we show that when the moral hazard cost of UI depends on local labor market conditions, this has important implications for the welfare consequences of UI. We develop a simple formula for the optimal level of unemployment benefits which takes into account how the behavioral response to UI benefits varies with local labor market conditions. The formula is stated in terms of our reduced-form parameter estimates and is thus in the spirit of the "sufficient statistics" approach to welfare analysis (Chetty 2009). The primary advantage of this method is that it can be implemented with relatively few parameter estimates.[6] Furthermore, these parameters can often be empirically estimated using a credible quasi-experimental research design. One disadvantage of this approach is that it is not well-suited to out-of-sample counterfactual analysis because the sufficient statistics are only valid for relatively "local" changes in the policy-relevant parameters. Using our reduced form empirical estimates to calibrate the optimal UI formula implied by our model, we find that a one standard deviation increase in the local unemployment rate leads

[6]We cannot conduct a full sufficient statistics analysis without reduced-form estimates of how the consumption smoothing benefits of UI vary with local labor market conditions. We hope that future work will build on Gruber (1997) and investigate this reduced-form effect.

to a 6.4 percentage point increase in the optimal replacement rate. To give a sense of the magnitude of this policy change, it is roughly equivalent to a one unit change in the coefficient of relative risk aversion in the model (e.g., from $\gamma = 2$ to $\gamma = 3$).

Several recent papers explore theoretically how UI benefits should vary with the unemployment rate (Kiley 2003, Costain and Reier 2005, Sanchez 2008 and Andersen and Svarer 2009). These papers differ in several respects. First, these papers take a structural approach to welfare analysis by imposing functional form assumptions characterizing how labor demand shocks affect search, while we take an approach in the spirit of the "sufficient statistics" literature, allowing us to use our reduced form estimates to calibrate our model. Second, our welfare analysis does not place any restrictions on the model primitives and is therefore valid for a wide range of underlying mechanisms which cause the duration elasticity to vary with unemployment. Third, these studies are primarily calibration analyses; they do not empirically estimate how the duration elasticity varies with local labor market conditions. Lastly, since these papers are mostly based on search models with no reservation wage decision, they do not highlight the distinction between the reservation wage and search effort elasticities.[7]

The remainder of the paper proceeds as follows. The next section develops the search model and describes both the agent and planner problems. Section 3 presents our empirical analysis which estimates how the behavioral response to UI varies with unemployment. Section 4 considers the welfare implications of our empirical findings. Section 5 concludes.

3.1 Theory

In this section, we describe the setup of a standard continuous-time, infinite-time horizon, job search model. The model closely follows Shimer and Werning (2007).

[7]This paper also contributes to a large empirical literature on the behavioral responses to UI by providing empirical evidence on how the elasticity of duration with respect to the benefit level varies with the unemployment rate. There are several papers in this area that indirectly relate to our work (Moffitt 1985, Arulampalam and Stewart 1995, Jurajda and Tannery 2003, and Røed and Zhang 2005).

We make a number of simplifying assumptions for tractability. First, we focus on benefit level, not potential benefit duration, although the latter is clearly an important policy parameter.[8] Second, the model does not allow workers to save or borrow. Thus an unemployed worker's only way to smooth consumption across states is the unemployment insurance agency.[9] Third, we omit leisure. Forth, we assume that workers are homogeneous. Finally, we work in a partial equilibrium setting with no firms. In ongoing work, we are working to relax each of these assumptions and evaluate the robustness of our results to these extensions. We begin by considering a version of the model where the job offer arrival rate is exogenous. We then extend the model to allow for endogenous search. In both cases, we characterize the structural relationship between the moral hazard cost of UI and unemployment. We then exploit this relationship to show how the welfare gain of UI varies with unemployment.

3.1.1 The Agent and Planner's Problems

Agent's Problem With Exogenous Arrival Rate. Consider a single worker that who has flow utility given by $U(c)$, where $U' > 0$, $U'' < 0$. The worker's subjective discount rate is given by $r \geq 0$. The worker maximizes the expected present value of utility from consumption

$$E\int_0^\infty e^{-rt}U(c(t))dt \tag{3.1}$$

If the worker is unemployed, she samples wages exogenously at rate λ from a known distribution function, $F(w)$. The distribution function possesses all of the properties that guarantee a solution exists. Workers who accept a wage offer commence employment immediately. Employment is assumed to end exogenously with separation rate s.

If the worker is unemployed, she receives and consumes an unemployment benefit

[8]Shimer and Werning (2007) find that socially optimal UI policy is infinite duration, constant benefits in both a hand-to-mouth model and one with free access to savings and lending.

[9]Since we assume that consumption during unemployment is equal to the UI benefit level and consumption during employment is equal to the net wage, there is full consumption-smoothing across time, within states.

denoted by b. When the worker is employed, she earns a wage w and pays taxes equal to τ which is used to finance unemployment benefit payments. Thus, her consumption is equal to her net wage, $w - \tau$.[10]

Finally, we assume that the model is stationary. Thus, λ, s, $F(w)$, b, τ and r are all assumed to be independent of time. The expressions that we derive in this paper depend on this assumption. For example, if there is duration dependence such that the reservation wage varies in response to the failure to find a job, then the expressions below will not be valid. Empirically, we do not find evidence of duration dependence in our data.

Worker Behavior. We now characterize worker behavior subject to a particular policy (b, τ). Let V_u be the value function (maximal expected lifetime utility) of an unemployed individual and let $V(w)$ denote the value function of a worker who accepts a wage offer of w. The workers solves the following:

$$rV_u = U(b) + \lambda \int_0^{\infty} \max\{V(w) - V_u, 0\} dF(w) \tag{3.2}$$

$$rV(w) = U(w - \tau) + s[V_u - V(w)] \tag{3.3}$$

where rV_u is the (per period) flow value of being unemployed, which is the consumption value plus the expected capital gain of getting an acceptable wage draw in the future (i.e., the "option value"). An employed worker earns $w - \tau$ and then at rate s loses her job and changes states, which she values at $V_u - V(w)$. Rearranging equation (3.3) results in the following expression:

$$V(w) = \frac{U(w - \tau) + sV_u}{r + s}$$

The reservation wage, w_R, satisfies $V(w_R) = V_u$, implying that $V(w_R) = U(w_R -$

[10]We do not model the worker's intensive labor supply decision. Since workers supply labor inelastically in our model, taxes are non-distortionary.

$\tau)/r$.[11] Substitution yields the following expression:

$$U(w_R - \tau) = U(b) + \frac{\lambda}{r+s}\int_{w_R}^{\infty}[U(w-\tau) - U(w_R-\tau)]dF(w) \tag{3.4}$$

Equation (3.4) is a standard expression in search models, which implicitly defines the reservation wage. The left-hand side of this equation represents the flow utility of accepting a wage offer of w_R. The right-hand side is the flow utility of rejecting a wage offer of w_R and waiting for a better wage draw. Note that $1/(r+s)$ represents the expected present value of a unit of income until a job ends. If there were no risk of job loss, this would be equal to $1/r$ which is the value of a perpetuity with payment of \$1. Therefore, the risk of job loss effectively increases the discount rate.

The job finding rate, p, is equal to the product of the job offer arrival rate and the probability of receiving an acceptable wage offer, $\lambda(1 - F(w_R))$. The stationarity assumption implies that p does not depend on how long the agent has been unemployed, meaning that we can express expected duration, D, as $1/p$.

Planner's Problem. We consider a social planner whose objective is to maximize an unemployed worker's utility, V_u. We restrict the class of feasible policies to those where the unemployment benefit level, b, and the employment tax, τ, are constant. We assume that the worker may receive UI benefits so long as she is unemployed. The planner's policy must satisfy a balanced-budget requirement which means that expected benefits paid out equals expected taxes collected, $Db = \frac{\tau}{r+s}$.[12] The right-hand side is roughly equal to the expected tax collected from the worker when she is employed. We solve the planner's problem in two steps: first, we show how the effect of UI on durations depends on unemployment; second, we exploit this relationship to show that the optimal benefit level chosen by the planner depends on the level of unemployment.

[11] Note that $V(w_R) = V_u \implies V(w) - V_u = \frac{U(w-\tau)-U(w_R-\tau)}{r+s}$. Also, $V(w_R) = \frac{U(w_R-\tau)+sV_u}{r+s} = \frac{U(w_R-\tau)+sV(w_R)}{r+s} = \frac{U(w_R-\tau)}{r}$.

[12] One may wonder why taxes are discounted, but unemployment benefits are not. This is because the government must pay benefits currently to a worker who is unemployed and receives taxes later, when the worker becomes employed.

3.1.2 Moral Hazard and Unemployment

In this reservation wage model, the moral hazard effect depends on responsiveness of reservation wages to benefits (Shimer and Werning 2007). This suggests evaluating this comparative static to see how the elasticity of duration with respect to the benefit level relates to the unemployment rate.

For simplicity, we start with the case where individuals are risk-neutral. Later, we will show how our main results generalize to the case of risk-averse workers. Define $u = \frac{s}{s+p}$ as the fraction of time a worker is unemployed or the unemployment rate.[13] The following lemma provides a simple expression for how the reservation wage responds to the benefit level.

Lemma 1 *For $r \approx 0$ and $U'' \approx 0$,*

$$\frac{\partial w_R}{\partial b} \approx u \tag{3.5}$$

This result is obtained by differentiating equation (3.4) with respect to the benefit level and applying Leibniz's rule for differentiation under an integral sign.[14] This expression is similar to the result obtained by Chesher and Lancaster (1983).[15] They were primarily interested in estimating the reservation wage and duration elasticities. In contrast, we are interested in using the search model to uncover the structural relationship between these elasticities and the unemployment rate.

There are several points worth making about expression (3.5). First, it implies we can measure the responsiveness of the reservation wage to changes in benefits in an extremely simple way – all that is needed is data on the unemployment rate.[16]

[13]Note that u and D have a 1-to-1 mapping in this model since $D = 1/p$ fully determines u, given s.

[14]Note that we will always slightly underestimate $\frac{\partial w_R}{\partial b} = \frac{r+s}{r+s+p} \equiv u^*$. The approximation error is likely to be small. To see this, note that when the unit of time is one month, $p \approx .46, s \approx .04$ (Shimer 2007) and $r \approx .004$ (Shimer and Werning 2007). Since r is about 1/10 the size of s in practice, the error $(u^* - u)/u^* = .09 * (1-u)$. Thus, the error is bounded above by 9%.

[15]In their model, $s = 0$ implying employment is an absorbing state. Since they do not observe p/r in their data, they express this in terms of the reservation wage, the conditional expected wage, and the unemployment benefit level. Specifically, $\frac{p}{r} = \frac{w_R - b}{x - w_R}$.

[16]Estimating the elasticity of the reservation wage with respect to the benefit level requires additional information on benefit levels and reservation wages, at a given unemployment rate.

For the U.S. over the period 1999-2009, $u \in [3.8\%, 10.1\%]$.[17] Feldstein and Poterba (1984) find empirically that $\frac{\partial w_R}{\partial b} \in [13\%, 42\%]$. As we show below, the marginal effect is higher when agents are risk-averse, which means that these estimates imply that risk aversion is relevant.

Second, each individual can be thought of as having her own "unemployment rate" since she optimally chooses her re-employment probability, p, to some extent. However, data limitations prevent us from calculating this expression at the individual level. Thus, in practice, we rely on the average unemployment rate across individuals.[18]

Third, note that we are not expressing the individual's decision problem explicitly in terms of the unemployment rate to see how it affects her behavior. This is different from decision problems that explicitly model the impact of aggregate variables on individual outcomes.[19] Rather, the result follows from the fact that the search model implies a steady-state relationship between the responsiveness of the reservation wage to benefits and the unemployment rate.

To see the intuition for this expression, let's consider the effect of a benefit increase. The key insight is that this increases benefits *in every period* that an agent remains unemployed. In a bad labor market, an agent is more likely to be unemployed for a long time, holding the reservation wage constant. Therefore, she stands to gain more at the margin from the increase in benefits than would be the case in a strong labor market where she is likely to become employed in the near future. Thus when benefits are increased and local labor market conditions are poor (i.e., local unemployment rate is high), an agent who is unemployed will need a higher wage to induce her into the workforce than would be the case for an agent who faces very favourable local labor market conditions.

Recognizing that w_R is a measure of the private welfare of the unemployed (since $V_u = V(w_R) \propto U(w_R - \tau)$), then an implication of this result is that a marginal

[17]Source: Bureau of Labor Statistics

[18]In ongoing work, we are constructing unemployment rates across observable demographic groups, using microdata from the CPS.

[19]For example, consider the consumer utility maximization problem where individual demand depends on the market price, which is determined in equilibrium.

increase in benefits increases the unemployed's private utility more when unemployment is bad. Note however, that from a social welfare perspective, what matters is the consumption smoothing benefit of UI which is positive only when the agent is risk-averse, as shown formally below.

We have shown that we can unambiguously determine how the responsiveness of reservation wages to benefits varies with unemployment. The following proposition considers how the duration elasticity varies with unemployment:

Proposition 2 *For* $r \approx 0$ *and* $U'' \approx 0$*,*

$$\varepsilon_{D,b} \approx \theta(w_R)ub \tag{3.6}$$

where $\theta(w_R) \equiv \frac{f(w_R)}{1-F(w_R)}$ *is the hazard rate (or failure rate) of the wage offer distribution.*

Proof. Differentiating $D = 1/p$ with respect to b yields the following:

$$\begin{aligned}\frac{\partial D}{\partial b} &= \frac{\lambda f(w_R)}{p}\frac{\partial w_R}{\partial b}D \\ &= \theta(w_R)\frac{\partial w_R}{\partial b}D \\ &\approx \theta(w_R)uD \end{aligned} \tag{3.7}$$

where the last line follows from Lemma 1. The result follows by multiplying $\frac{\partial D}{\partial b}$ by b and dividing by D. ■

This expression is positive so that an increase in b raises w_R and increases D. The fact that benefits increase unemployment does not necessarily mean the individual is worse off. Since she *chooses* to be unemployed longer, by revealed preference, she must be better off from a private welfare standpoint.[20] Expression (3.6) shows that the duration elasticity depends on three factors: (1) the hazard rate of the wage offer distribution, (2) the unemployment rate and (3) the unemployment benefit level.

[20]This does not imply that social welfare is increased since the agent imposes a negative externality on the government's budget. We return to the normative implications below.

How the duration elasticity varies with the unemployment rate depends crucially on how $\theta(w_R)$ varies with u.

We assume that F does not vary directly with the unemployment rate.[21] In order to sign this effect, we need to know how w_R varies with u and how $\theta(w_R)$ varies with w_R.[22] Consider the relationship between w_R and u. The first thing to recognize is that w_R and u are jointly determined and therefore are not causally related.[23] This implies that their relationship will in general depend on the underlying sources of variation.[24]

Because $\frac{\partial \theta(w_R)}{\partial w_R}$ depends on the shape of the wage distribution, the relationship between the unemployment rate and the duration elasticity in a reservation wage model is theoretically ambiguous. According to Van den Berg (1994), most of the distributions used in structural job search analysis have hazards that are decreasing in w_R, $\frac{\partial \theta(w_R)}{\partial w_R} < 0$. In that case, then the model unambiguously predicts that the moral hazard cost of UI increases during recessions, in contrast to the hypothesis of Krueger and Meyer discussed in the introduction. We have also calibrated the model when wages are distributed log-normally, and we also find that the duration elasticity is positively related to the unemployment rate.[25]

Since the job offer arrival rate is exogenous, the relationship between the unem-

[21]This assumption is consistent with the large macroeconomics literature that provides evidence showing that wages are acyclical (Bewley 1999).

[22]Chesher and Lancaster (1983) show that when the wage offer distribution for $w \geq b$ is Pareto, $\theta'(w_R) < 0$. On the other hand, when it is Normal, $\theta'(w_R) > 0$. More generally, any distribution that is log-concave will have a non-decreasing hazard function (see Burdett 1981).

[23]This casts doubt on research that empirically estimates the relationship between unemployment or unemployment duration and reservation wages. As Jones (1988) points out, job search theory implies that most variables influencing employment probabilities given the reservation wage can be expected to also influence the reservation wage and as a result, the exclusion restriction is likely to be violated.

[24]As an analogy, consider the relationship between the equilibrium values of price and output. A positive demand shock increases price and output, so that the two variables are positively correlated. A negative supply shock on the other hand, increases price and lowers output, so that the two variables are negatively correlated. Variation in unemployment driven by local labor market conditions (e.g., variation in λ) will cause the responsiveness of the duration elasticity to local labor market conditions to depend on $\frac{\partial w_R}{\partial \lambda}$. In this model, $\frac{\partial w_R}{\partial \lambda} > 0$.

[25]The log-normal distribution does not have a monotonic hazard rate, so we cannot sign the association between the duration elasticity and the unemployment rate analytically. However, for a large range of plausible parameter values, we consistently found that the association was positive, just as when the wage offer distribution was Pareto.

ployment rate and the duration elasticity is determined solely through changes in reservation wage response. Below we consider a more realstic model where workers also choose search effort to affect the job offer arrival rate. Before we turn to this richer model, we briefly consider the case where workers are risk-averse, and we carry through the assumption that workers are risk-averse for the remainder of this section.

Risk Aversion

With risk aversion, it can be shown that for small values of r

$$\frac{\partial w_R}{\partial b} \approx \frac{U'(b)}{U'(w_R - \tau)} u \tag{3.8}$$

Relative to the risk-neutral case, the marginal effect is amplified by the ratio of marginal utilities since $b < w_R - \tau$ and $U'' < 0$. Intuitively, a risk-averse agent values a guaranteed stream of unemployment benefits more than a risk-neutral agent and so is more sensitive to variations in her certain income. This also implies that

$$\begin{aligned} \frac{\partial D}{\partial b} &= \theta(w_R) \frac{\partial w_R}{\partial b} D \\ &\approx \frac{U'(b)}{U'(w_R - \tau)} \theta(w_R) u D \end{aligned} \tag{3.9}$$

Therefore,

$$\varepsilon_{D,b} \approx \frac{U'(b)}{U'(w_R - \tau)} \theta(w_R) u b \tag{3.10}$$

Thus, relative to the risk-neutral case, the duration elasticity is amplified by the ratio of the marginal utilities when unemployed and employed, respectively.

Incorporating Endogenous Search

The search model shows that UI benefits raise unemployment durations since they put upward pressure on reservation wages, which in turn reduces the probability that a worker gets an acceptable wage offer. Some empirical studies, however, have found that increases in benefits do not affect the distribution of accepted wage offers, implying that the effect on reservation wages is small (see Card, Chetty, and Weber

2007).[26] In this section, we allow for the possibility that individuals can affect the job offer arrival rate through costly search effort (Rogerson, Shimer, and Wright 2005). This provides an additional channel through which UI benefits can increase the length of unemployment spells.

Let search effort be denoted by e and let the arrival rate be given by $\lambda(e)$, where $\lambda' \geq 0$ and $\lambda'' \leq 0$. In this case, it can be shown that

$$\varepsilon_{D,b} = \theta(w_R)\frac{\partial w_R}{\partial b}b - \delta(e)\frac{\partial e}{\partial b}b \tag{3.11}$$

where $\delta(e) \equiv \frac{\lambda'(e)}{\lambda(e)}$. Clearly, how the duration elasticity varies with unemployment depends crucially on how $\frac{\partial e}{\partial b}$ varies with u in addition to how $\frac{\partial w_R}{\partial b}$ varies with u. Thus, adding search intensity to the model potentially changes how moral hazard varies unemployment. The first part of expression (3.11) is simply the duration elasticity with no search decision. The second term of expression (3.11) shows that the duration elasticity with search depends on how UI benefits distort search effort $(\frac{\partial e}{\partial b})$ as well as on how the arrival rate varies with search effort $(\delta(e))$. The key "behavioral" parameters of this expression are $\frac{\partial w_R}{\partial b}$ and $\frac{\partial e}{\partial b}$; the terms $\theta(w_R)$ and $\delta(e)$ primarily depend on the economic environment and are only indirectly affected by the behavioral effects.

To analyze the marginal effects in this expression, we need to study the optimality conditions for search and the reservation wage. We assume that the search cost, denoted by $\psi(e)$, is strictly increasing and convex and is separable from consumption utility.[27] To simplify the algebra, it will be convenient to define the *surplus function* $\varphi(w_R) \equiv \int_{w_R}^{\infty}[U(w-\tau) - U(w_R-\tau)]dF(w)$. This represents the difference between the optimized values of employment and unemployment. Intuitively, it measures a worker's expected utility when employed relative to her "reservation employment

[26] One can show that the expected wage satisfies $E_w[w|w \geq w_R] = w_R + \frac{\int_{w_R}^{\infty}[1-F(w)]dw}{1-F(w_R)}$. Thus, if benefits do not affect average wages, they must not affect reservation wages.

[27] While the separability between search and consumption simplifies the analytics, it may be the case that unemployed individuals affect the job offer arrival rate by changing consumption. This formulation would affect the welfare analysis. In ongoing work, we are studying the effects of incorporating this generalization.

utility" – the utility she receives at the wage she is just willing to accept to become employed.

We will use the following property of the surplus function, $\frac{\partial \varphi(w_R)}{\partial w_R} = -(1 - F(w_R))U'(w_R - \tau)$. This is negative since holding the distribution of wages fixed, a worker gets less surplus when her reservation utility is higher. We show below that $\frac{\partial w_R}{\partial b} > 0$ implying $\frac{\partial \varphi(w_R)}{\partial b} < 0$. Intuitively, an increase in unemployment benefits raises the value of unemployment and hence the reservation wage, and in turn lowers the expected net surplus from employment. In a consumer demand setting, this would be represented by an inward shift of the demand curve which lowers consumer surplus for a fixed market price. Note that one can also show that $\frac{\partial \varphi(w_R)}{\partial \lambda} < 0$. Intuitively, a higher arrival rate increases the option value of unemployment. The implicit equation for the reservation wage can be written compactly as

$$U(w_R - \tau) = U(b) - \psi(e) + \frac{\lambda(e)}{r+s}\varphi(w_R)$$

The optimal e can be found by maximizing $U(w_R - \tau)$. The first-order condition assuming an interior optimum is

$$\psi'(e) = \frac{\lambda'(e)}{r+s}\varphi(w_R) \tag{3.12}$$

Thus, the optimal search level equates the marginal cost of effort (left-hand side) with the marginal value of effort (right-hand side). The marginal value of effort depends on the marginal increase in the likelihood of obtaining a job in response to an increase in effort and the expected discounted surplus of getting a job. Note that searching harder only affects the likelihood of getting an offer, but does not affect expected income, conditional on getting a job.[28] The model therefore predicts that a positive shift in local labor demand increases search intensity of the unemployed. In other words, search intensity in this model is negatively correlated with unemployment.

[28]This follows from the assumption that search effort affects only the arrival rate, not the wage distribution.

Substituting this equation into the reservation wage equation yields the following expression:

$$U(w_R - \tau) = U(b) + \frac{\lambda(e)}{\lambda'(e)}\psi'(e) - \psi(e) \tag{3.13}$$

The conditions (3.12) and (3.13) comprise a system of equations, which implicitly (and jointly) determine the optimal reservation wage and the optimal level of search effort, as functions of the level of UI benefits. We can differentiate this system with respect to b to solve for $\frac{\partial w_R}{\partial b}$ and $\frac{\partial e}{\partial b}$.

Proposition 3 *Assume r is small. The marginal effects with endogenous search intensity satisfy*

$$\frac{\partial w_R}{\partial b} = \frac{U'(b)}{U'(w_R - \tau)}u \tag{3.14}$$

$$\frac{\partial e}{\partial b} = -\frac{\delta(e)(1-u)U'(b)}{\psi''(e) - \frac{\lambda''(e)}{s}\varphi(w_R)} \tag{3.15}$$

Proof. See Appendix A. ■

First, consider the expression for $\frac{\partial w_R}{\partial b}$. Note that adding endogenous search effort does not change the formula for how the reservation wage responds to the benefit level. However, $\frac{\partial w_R}{\partial b}$ still depends on search effort indirectly through u and w_R. Next, consider the expression for $\frac{\partial e}{\partial b}$. Note that $\frac{\partial e}{\partial b} < 0$; that is, an increase in benefits lowers the marginal gain of search since it decreases expected surplus from employment, $\varphi(w_R)$. The magnitude of this decrease in search effort is determined by three factors: (1) the initial shift in the marginal benefit curve $(-\delta(e)(1-u)U'(b))$, (2) the slope of the marginal cost curve (ψ'') and (3) the slope of the marginal benefit curve $(\frac{\lambda''(e)}{s}\varphi(w_R))$.

To interpret the shift in the marginal benefit curve, recall that the marginal benefit curve as a function of effort is given by $\frac{\lambda'(e)}{r+s}\varphi(w_R)$. Therefore, the shift in response to a change in the level of benefits, for a fixed level of effort, depends on the magnitude of $\frac{\lambda'(e)}{r+s}$ and also on how $\varphi(w_R)$ responds to a change in benefits. The first term which relates to $\delta(e)$ illustrates that the location of the curve matters for the size of the shift. Intuitively, a small value for $\delta(e)$ implies that the arrival rate does not respond much to a marginal increase in effort, lowering the level of the marginal benefit curve and essentially placing a bound on how distortionary benefits can be. To see why the employment rate $1-u$ matters, consider the case where $s \to \infty$, so that $u=1$. In this case, there is no chance of actually being employed so workers essentially put no weight on expected surplus from employment. That the shift depends on $U'(b)$ follows since this term characterizes how $\varphi(w_R)$ responds to the benefit level.

That $\frac{\partial e}{\partial b}$ depends on the slopes of the curves follows from any standard marginal analysis. If ψ'' is large at a given level of search effort, this means that marginal cost curve is inelastic. As a result, a given reduction in the marginal benefit curve due to an increase in benefits has less of an impact on search effort. Similarly, the effort response depends on the slope of the marginal benefit curve, which is pinned down by λ''. A small value for λ'' implies that the marginal benefit curve is more elastic; hence a reduction in benefits have a larger effect on search.

Examining expression (3.15), we can see that a decline in local labor demand can impact the distortionary effect of UI benefits on search through it's effect on $\delta(e)(1-u)$.[29] Clearly, a negative labor demand shock is going to increase the unemployment rate, so the sign of the effect ultimately depends on how the shock affects $\delta(e)$. As a reminder, $\delta(e)$ represents the *percentage change* in the job offer arrival rate from an additional unit of search. A larger value of $\delta(e)$ means that search is more productive. Thus, the key determinant of the comparative static is whether search is more productive on the margin in a weak or strong local labor market. In a weak market, we would expect $\lambda(e)$ to be small, which would act to increase $\delta(e)$. On the other hand, $\lambda'(e)$ is also likely to be smaller which lowers $\delta(e)$, so the net effect depends on the rate at which $\lambda'(e)$ falls relative to $\lambda(e)$. As Kiley (2003) discusses, it is possible to specify functional forms so that the net effect can go either way. As a result, the question is ultimately an empirical one. We calibrate the model below by assuming a particular functional form, $\lambda(e) = \overline{\lambda} + \lambda e$. Here $\lambda'(e)$ falls at rate 1 with λ and $\lambda(e)$ falls at rate $e + \lambda \frac{\partial e}{\partial \lambda}$. So, the net effect depends on whether $e(1 + \frac{\lambda}{e}\frac{\partial e}{\partial \lambda}) > 1$. With this functional form assumption, $\lambda'' = 0$ giving

[29]It is possible that the recession affects the term $\frac{\lambda''(e)}{s}\varphi(w_R)$, although signing this effect seems less intuitive.

$$\frac{\partial e}{\partial b} = -\frac{U'(b)}{\psi''(e)}\delta(e)(1-u)$$

To see what this implies about the duration elasticity, let's plug the marginal effects into the elasticity formula[30]:

$$\varepsilon_{D,b} = \frac{U'(b)}{U'(w_R - \tau)}\theta(w_R)ub + \frac{U'(b)}{\psi''(e)}(\delta(e))^2(1-u)b \tag{3.16}$$

This expression shows that whether moral hazard increases or decreases in a recession depends on the relative strength of the reservation channel and search channel. We present a calibration in the next section that is an attempt to disentangle these two channels and also show independently how they vary with the unemployment rate.[31]

3.1.3 Calibrating $\varepsilon_{D,b}$

Our expression for $\varepsilon_{D,b}$ demonstrates there are two channels by which unemployment can impact moral hazard. This section evaluates the duration elasticity numerically by calibrating the model in the previous section. The calibration sheds light on the plausible quantitative impact of the local unemployment rate on the duration elasticity before turning to the empirical results.

Functional Form Assumptions

In what follows, we rely on Chesher and Lancaster (1983), Shimer (2007) and Chetty (2008). A unit of time for the calibrations is a week. For all of these calculations, we assume $r = 0$.

Wage Offer Distribution. We assume wages are distributed log-normally, with mean weekly wages of \$300 and the standard deviation of weekly wages is \$240. In

[30]Note that this is the partial elasticity, which captures the effect of a change in benefits on expected duration, holding taxes constant.

[31]In future work, we will allow the distribution of wages to also vary with local labor demand conditions.

the Appendix, we follow Chesher and Lancaster (1983) and assume that the wage offer distribution is Pareto.

Arrival Rates. We assume that the arrival rate takes a linear form, $\lambda(e) = \overline{\lambda} + \lambda e$. Separations end exogenously at rate s. Shimer (2007) reports estimates for the job finding and separation probabilities. There is a simple connection between the rates and probabilities. To see this, note that the probability that a worker has not found a job after a spell of length t is $P \doteq e^{-Ht}$, where $H = \lambda(1 - F(w_R))$. Therefore the job finding probability is $1 - P = e^{-(1-H)t}$. It follows that the job finding rate is $-\log(1 - P)$. There is a similar connection between the separation probability and the separation rate. Shimer (2007) finds that the average monthly separation probability in the US from 1948 to 2004 is 0.035. This delivers a separation rate of .02. Converting this to a weekly rate yields $s = .00387$.

UI Benefits. Assume benefits are equal to $b = r \times E[w]$. Following Chetty (2008), we take $r = 0.5$. For these simulations, we assume no taxes, so $\tau = 0$.

Preferences over consumption. Assume that $U(c) = \frac{c^{1-\gamma}}{1-\gamma}$, where $\gamma > 0$ is the risk aversion parameter (As $\gamma \to 1$, $U(c) \to \log c$). We follow Chetty (2008) by choosing $\gamma = 1.75$.

Search Effort. We let search costs as a function of effort be denoted by $\psi(e) = \phi \frac{e^{1+\kappa}}{1+\kappa}$, where ϕ is a scaling parameter. The elasticity of search costs with respect to search effort is $1 + \kappa$. So a higher κ increases the marginal cost of search and lowers search effort.

Results

Tables 1 and 2 show the results from our calibration. The experiment we consider is to exogenously vary the term λ in the fuction $\lambda(e) = \overline{\lambda} + \lambda e$. Each column in the table represents a different value of λ. To shed some light on the underlying mechanisms, we report the total duration elasticity in equation (3.16) as well as each of the two terms that comprise the duration elasticity, separately.

In Table 1 we choose a low value of $\overline{\lambda}$ (= 0.02). Looking across the second row ($\varepsilon^{w_R}_{D,b}$), it is clear that an increase in λ increases the responsiveness of reservation wages to UI benefits. The third row ($\varepsilon^{e}_{D,b}$) shows that an increase in λ increases the responsiveness of search effort to UI benefits. Since both move in the same direction, the duration elasticity also increases as λ increases, causing the duration elasticity to be increasing in the unemployment rate. As a result, in this calibration, the moral hazard cost of UI *increases* with the unemployment rate.

Table 2 reports results choosing a higher value of $\overline{\lambda}$ (= 0.1). In this table, we can see looking across the second row ($\varepsilon^{w_R}_{D,b}$) that an increase in λ increases the responsiveness of reservation wages to UI benefits, just as with Table 1. However, unlike Table 1, the third row ($\varepsilon^{e}_{D,b}$) of Table 2 shows that an increase in λ *decreases* the responsiveness of search effort to UI benefits. In this calibration, the search effort effect dominates the reservation wage effect, so that the duration elasticity also decreases as λ increases, causing the duration elasticity to be decreasing in the unemployment rate. As a result, in this calibration, moral hazard *decreases* with the unemployment rate.

This analysis demonstrates the importance of incorporating endogenous search intensity, and that the precise way in which local labor market conditions affect the returns to search effort ultimately determines whether moral hazard increases or decreases with the unemployment rate.

We now turn to what these results imply for optimal policy. In practice, to examine how moral hazard varies with unemployment, we do not need to separately identify the responsiveness of the reservation wage and search intensity to benefits; we only need to identify the duration elasticity. This motivates our empirical strategy, described later, which explores how the duration elasticity varies with the unemployment rate. The next section presents a welfare analysis to show how the optimal benefit level varies with the local labor market conditions.

3.1.4 Welfare Analysis: Optimal Unemployment Benefits

The social planner solves the following problem

$$\max_{b,\tau} V_u$$

$$s.t.\ Db = \frac{\tau}{r+s}$$

Since $V_u = U(w_R - \tau)/r$, the planner's problem is simply to maximize the worker's after-tax reservation wage, $w_R - \tau$. The following theorem characterizes the optimal benefit level.

Theorem 4 *The optimal benefit level satisfies the following condition:*

$$\frac{U'(b) - U'(w_R - \tau)}{U'(w_R - \tau)} = \varepsilon_{D,b} \tag{3.17}$$

Proof. See Appendix A. ∎

An instructive derivation of this first-order condition is as follows.[32] Let $\overline{g} = \frac{U'(b)}{U'(w_R - \tau)}$ be the amount such that, the government is indifferent between giving \$1 to someone who is unemployed and $\overline{g}$ to someone who is employed. Next, consider a \$1 increase in benefits. This has a mechanical effect on UI expenditures and a behavioral response. The mechanical effect, M, is given by Ddb. The change in expenditures due to behavioral responses, B, is given by $\frac{\partial D}{\partial b} bdb = \varepsilon_{D,b} \frac{D}{b} db$. To obtain the optimal benefit level, we must equalize the expenditure effect, $M + B$, with the welfare effect. A simple application of the envelope theorem implies that the welfare effect is given by $\overline{g}M$. That is, each additional dollar raised by the government to finance UI benefits reduces on average social welfare of the employed by $\overline{g}M$. Thus, at an optimum, $(1 - \overline{g})M + B = 0$. Rearranging this equation delivers the result.

The test for the optimality of UI benefits compares the difference in consumption between an unemployed worker and a worker employed at her reservation wage with the moral hazard cost of social insurance. This is slightly different than the consumption-based test in Chetty (2006) as the consumption smoothing measure here corresponds to the difference between the lowest acceptable level of consumption while employed and consumption while unemployed, rather than the difference between average consumption while employed and unemployed. The reason for this difference is due to the difference in the maximand in the social planner's problem.

[32] This derivation closely follows the derivation of the optimal top tax rate in Saez (2001).

In Chetty (2006), the maximand is expected utility. Thus, the social planner trades off consumption utility between the states of employment and unemployment. In this model, the maximand is an unemployed worker's utility. The social planner trades off current consumption utility with the change in utility (surplus) that occurs if the worker becomes employed.

Two final points are worth mentioning. First, since consumption when employed exceeds the net reservation wage, the optimal benefit level in this setting is lower than the optimal benefit level if the planner was interested in maximizing expected utility. The reason is because insurance is more valuable when the state of nature has yet to be realized. Finally, in practice, $\varepsilon_{D,b}$ will not be zero, so we will have $b < w_R - \tau$.

The Optimal Benefit Level and Unemployment

To see how the optimal benefit level varies with the unemployment rate, we need to consider how both sides of equation (3.17) vary with unemployment. Since we already considered how the moral hazard cost of UI varies with unemployment, let us focus our attention on how the consumption smoothing or insurance effect varies with unemployment. Unemployment has an effect on the left-hand side of equation (3.17) that operates through the balanced-budget constraint. To see this, consider the case where UI benefits are not distortionary. The government budget constraint implies that for $r = 0$,

$$\frac{\partial \tau}{\partial b} = \frac{u}{1-u}$$

Thus, when unemployment is high, more taxes need to be raised to finance a given level of benefits. This shows that the insurance effect depends indirectly on labor market conditions. In particular, this implies that benefits should fall when unemployment increases. To see this, note when unemployment increases for a given level of benefits, to satisfy the balanced-budget condition, taxes must increase on the employed. This lowers the marginal utility of consumption for the employed relative to marginal utility of consumption for the unemployed (e.g., $w_R - \tau$ is reduced); in order to restore optimality, benefits need to be reduced. Andersen and Svarer (2009) label this a "budget effect" since the effect comes purely from the need to satisfy the budget constraint.

Let us consider how the optimal benefit level b^* varies with the job offer arrival rate $\lambda(e)$. This is given in the following collary.

Corollary 5 *The effect of a change in the offer arrival rate on the optimal benefit level is given by*

$$\frac{\partial b^*}{\partial \lambda} = \frac{U''(w_R - \tau)\frac{\partial (w_R - \tau)}{\partial \lambda}\varepsilon_{D,b} + U'(w_R - \tau)\frac{\partial \varepsilon_{D,b}}{\partial \lambda}}{U''(b)} \tag{3.18}$$

The proof follows from differentiating condition (3.17) with respect to $\lambda(e)$. The first term in expression (3.18) represents the budget effect. Since $U'' < 0$ and $\frac{\partial (w_R - \tau)}{\partial \lambda} > 0$, the budget effect causes benefits to be lower when unemployment is higher. The second term is the effect on distortions. If the moral hazard effect of UI increases with the job finding rate (when unemployment is low), this term causes benefits to be higher when unemployment is high. Thus, the net effect of unemployment on the optimal benefit level depends on the relative strengths of the budget effect and the distortion effect. Once we have empirical estimates of how the duration elasticity varies with local labor market conditions, we can use the estimates to calibrate the social planner's optimal UI problem to compute how UI benefit levels should optimal respond to local labor market conditions. The next section describes our empirical strategy which estimates how the duration elasticity varies with unemployment.

3.2 Estimation Strategy and Data

Our empirical strategy consists of two parts: (1) graphical evidence and nonparametric tests of survival curves and (2) semi-parametric estimates of proportional hazard models (Cox models). The empirical strategy closely follows Chetty (2008).

We use unemployment spell data from the SIPP spanning 1985-2000. We impose the same restrictions as in Chetty (2008): we focus on prime-age males who (a) report searching for a job, (b) are not on temporary layoff, (c) have at least three months of work history, and (d) took up UI benefits. We focus on two alternative proxies for individual's actual UI benefits: (1) average benefits for each state-year pair and (2) maximum weekly benefit amount. In ongoing work, we are working to implement an instrumental variables hazard model, where the goal is to construct a simulated instrument which isolates policy variation in individual UI benefits that is driven purely by change in UI laws (Gruber 1997).

3.2.1 Graphical evidence and nonparametric tests

We begin by providing graphical evidence on the effect of unemployment benefits on durations. We split the sample into two sub-samples, according to whether individuals begin their unemployment spell in states with above-median unemployment or in states with below-median unemployment. Each year we define the median unemployment rate across states. We categorize a state as having either above or below median unemployment that year. We assign unemployment ratess to unemployment spells based on the unemployment rate in the state that the individual resides in when the spell began, using monthly data on state unemployment rates. We also categorize unemployment spells based on whether the UI benefit level in a given state and year is above or below the median UI benefit level for that year.

Figures 1 and 2 show the effect of UI benefits on the probability of unemployment for individuals in above-average and below-average unemployment state-years, respectively. In each figure, we plot Kaplan-Meier survival curves for individuals in low-benefit and high-benefit states.[33] The results in figure 1 show that the curves are fairly similar in both low-benefit and high-benefit states when the unemployment rate in a state-year is above the median unemployment rate. The curve in high-benefit states is slightly higher, indicating that UI benefits may marginally increase benefits, but a nonparametric test that the curves are identical does not reject at conventional

[33]Following Chetty (2008), the plotted curves are adjusted for the "seam effect" in the SIPP panel data, but the test that the survival curves are identical is fully nonparametric and does not make this adjustment.

levels ($p = 0.156$). By contrast, in figure 2 the curves are noticeably different; in particular, the durations are significantly longer in high-benefit states, and the difference between the survival curves is strongly statistically significant ($p < 0.001$).

These figures show that the moral hazard effect of UI benefits depends crucially on whether unemployment is high or low. In particular, our findings suggest that the effect of UI benefits on durations is not statistically significant when the unemployment rate is high but is strongly statistically significant when the unemployment rate is low.[34] These comparisons are based on simple comparisons across spells. It is possible, however, that the characteristics of individuals vary with unemployment rate in a way that would bias these comparisons. To investigate this potential bias, the next subsection reports semi-parametric proportional hazard models which include a rich set of individual-level controls. The results from the hazard models are broadly consistent with the results based on these figures.

3.2.2 Semiparametric Hazard Models

We investigate robustness of graphical results by estimating a set of Cox proportional hazard models in Tables 4 through 8.[35] Each table reports results with alternative sets of control variables in the columns. The baseline estimating equation is the following:

$$\log d_{i,s,t} = \alpha_t + \alpha_s + \beta_1 \log(b_{i,s,t}) + \beta_2(\log(b_{i,s,t}) \times u_{s,t}) + \beta_3 u_{s,t} + \mathbf{X}_{i,s,t}\Gamma + e_{i,s,t} \quad (3.19)$$

where $d_{i,s,t}$ is the duration of the unemployment spell, α_t and α_s represents year and state fixed effects, $b_{i,s,t}$ is the unemployment benefit for individual i at start of spell, $u_{s,t}$ is the state unemployment rate at the start of the spell and $\mathbf{X}_{i,s,t}$ is a set of (possi-

[34]We have also looked at the subsample of workers with above-median liquid wealth, and we find broadly similar results (see Appendix Figures A1 and A2). These results suggest that liquidity effects are not primarily accounting for the differential duration elasticity between high and low unemployment, which is broadly consistent with our results in Table 7, described below.

[35]We are looking into alternative semiparametric hazard models to broaden the scope of the empirical analysis. Concerns have been raised that Cox models may not be reliable in the presence of ties. As such, we are going to report Han-Hausman estimates which are more reliable when the number of ties is large relative to the sample size.

bly time-varying) control variables.[36] The unemployment rate at the start of the spell is de-meaned so that the coefficient β_1 gives the elasticity of unemployment durations with respect to UI benefits at average levels of unemployment. The coefficient on the interaction term (β_2) gives the incremental change in the duration elasticity for a one percentage point change in the state unemployment rate.

Before turning to our regression results, we present descriptive statistics in Table 3. The table presents summary statistics for the overall sample and the two sub-samples used to create figures 1 and 2. One can see that in high unemployment states, average income, education, the fraction married, UI benefits, are all lower than in low unemployment states. Individuals are also slightly older in these states. Since the distribution of observables is different across the two samples, one question that arises when considering how the duration elasticity varies with unemployment is whether this relationship is coming from "selection" (i.e., compositional changes in the unemployed population due to changes in the local labor market conditions) and how much of it is coming from an actual change in the behavioral response. This will depend on the extent to which the duration elasticity varies directly with demographics, which we investigate in detail in Table 6 below.

The main results are reported in Table 4. Column (1) of Table 4 reports results of a specification broadly similar to the previous literature (Moffitt (1985), Meyer (1990), Chetty (2008).[37] This specification controls for age, marital status, years of education, a full set of state, year, industry and occupation fixed effects, and a 10-knot linear spline in log annual wage income. The results indicate that the elasticity of durations with respect to the UI benefit level is -0.651 (s.e. 0.318) and the estimate is statistically significant at conventional levels ($p = 0.041$). Column

[36]The notation of the estimating equation is a simplified presentation of the actual model. The actual (latent) hazard rate is the true left-hand side variable, but is not actually observed in the data; additionaly, there is a flexible (nonparametric) baseline hazard rate which is also estimated when fitting the Cox proportional hazard model. Following Chetty (2008), we fit a separate baseline hazard rate for each quartile of net liquid wealth, although our results are similar when a single nonparametric baseline hazard rate is estimated instead.

[37]The results are not identical to Chetty (2008) because of slightly different sample restrictions, the inclusion of the state uemployment rate as an additional control, and becauase we estimate a more flexible baseline hazard function (where we nonparametrically estimate a separate baseline hazard for each quartile of net liquid wealth).

(2) reports estimates of equation (3.19) above. This column includes the same set of controls in column (1) and estimates the same hazard model; the only difference is the addition of an interaction term between the UI benefit level and the state unemployment rate. The coefficient on the interaction term (β_2) represents the change in the duration elasticity for a one percentage point increase in the state unemployment rate. The results in column (2) show an estimate of β_2 of 0.142 (s.e. 0.068). The bottom two rows show an alternative way to interpret the interaction term. These rows report the duration elasticity and one standard deviation above and below the mean unemployment rate. At one standard deviation below the mean, the duration elasticity is 0.502 (s.e. 0.326), while at one standard deviation above the mean the duration elasticity is 0.980 (s.e. 0.388). These results imply that the moral hazard effect of UI varies significantly with unemployment, and that the magnitude of the duration elasticity is decreasing with local labor market conditions.

Robustness Tests

Alternative Measures of Interaction Term. Table 5 reports results which replace the interaction of UI benefit generosity (average weekly benefit amount) and the state unemployment rate with alternative measures of each variable in the interaction term. Each row reports alternative measures of the interaction term.

The first row of Table 5 reproduces our baseline estimates for comparison. The second row replaces the state unemployment rate with a dummy for whether or not the unemployment rate is greater than the median state unemployment rate in that year. This specification corresponds more closely to the nonparametric results presented above. The third row replaces the average weekly benefit amount with the maximum weekly benefit amount. The maximum weekly benefit amount corresponds more to a specific policy parameter that states directly adjust from time-to-time. Thus, the robustness of the estimates to the use of this measure is likely to shed some light on whether the variation in average weekly UI benefits is plausibly exogenous (conditional on state and year fixed effects). The estimates of the interaction term is similar in magnitude to the baseline specification.

Finally, in the last three rows, we report results that are based on two separate measures of the unemployment rate. The search model predicts that variation in the unemployment rate due to an increase in the job separation rate should have a similar effect on the duration elasticity as a reduction in the job finding rate. To test this hypothesis, we follow the methodology proposed in Shimer (2007) which estimates the job finding and job separation rates, based on gross unemployment and employment flows. Shimer shows that the job finding probability F_t satisfies the following equation:

$$u_{t+1} = (1 - F_t)u_t + u^s_{t+1}$$
$$F_t = 1 - \frac{u_{t+1} - u^s_{t+1}}{u_t} \tag{3.20}$$

where u_t is the number of unemployed workers in period t and u^s_{t+1} is the number of unemployed workers at the end of the period who were employed at some point during the period. Thus, with data on the unemployed, we can construct a measure of the job finding probability and job finding rate, $f_t \equiv -\log(1 - F_t)$.[38] Next, Shimer shows that the job separation rate, s_t, satisfies

$$u_{t+1} = \frac{(1 - e^{-f_t - s_t})s_t}{f_t + s_t} l_t + e^{-f_t - s_t} u_t$$

where $l_t = u_t + e_t$ and e_t is the number of employed workers in period t. Given the empirical measures f_t and s_t, we construct the following two "unemployment rates":

$$u_f = \frac{\overline{s}}{\overline{s} + f_t}$$
$$u_s = \frac{s_t}{s_t + \overline{f}}$$

where the bar means that they are average values during the sample period. The unemployment rates u_f, u_s measure variation in unemployment coming purely from

[38]In practice, data on u^s_{t+1} by state is not publicly available. Thus, we make a simplifying assumption by assuming that $\frac{u^s_{t+1}}{u_t}$ is identical across states. In ongoing work, we are constructing short-term unemployment using microdata from the CPS.

variation in f_t and s_t, respectively. Rows (4) and (5) in Table 5 report results from interacting benefits with these two measures separately, and row (6) reports results from including both measures together. This allows us to see separately how variation in unemployment rate coming from the separation rate and the job finding rate affect the duration elasticity. Interestingly, we find that the effects in the two rows are fairly similar to the baseline specification and also fairly similar to each other (the p-value of the test that the two interaction terms in row 6 are equal is 0.731). Thus we conclude that the variation in unemployment rate affects the duration elasticity regardless of whether that variation is coming from the job finding rate or the job separation rate.

Composition Bias and Selection on Observables. As explained above, the observation that the duration elasticity varies with unemployment can in principle be explained by two possibilities: first, a change in a given individual's job finding or job separation rate directly changes her responsiveness to benefits. Alternatively, if there is heterogeneity in moral hazard across demographic groups and the distribution of demographics of the unemployed varies with the level of unemployment, then this compositional change could be responsible for the change in the average duration elasticity. To test how much of the magnitude is coming through this compositional channel, we report estimates of our baseline specification where we add interactions between benefits and the demographic controls in the baseline specification: age, marital dummy, years of education, occupation fixed effects, and industry fixed effects. If the estimates of the interaction term in the baseline specification is mostly due to compositional changes (among demographic groups with different duration elasticities), then we would expect to see a reduction in the magnitude of the coefficient on the interaction between benefits and unemployment. Table 6 shows that our main result is quite robust to including such controls. Looking across columns, we see that adding interactions between demographics and benefits does not change the coefficient on our main coefficient of interest (the interaction term) in any substantive way. This appears to be primarily due to the fact that the duration elasticity does not appear to vary greatly with observable demographics.[39]

Moral Hazard versus Liquidity. Recent work by Chetty (2008) raises a concern with interpreting the duration elasticity as a pure moral hazard effect. He presents compelling evidence that part of the observed duration elasticity is due to a "liquidity effect." This suggests that the interaction term which we estimate in our baseline specification could plausibly represent a liquidity effect which varies systematically with local labor market conditions. We deal with this concern in two ways. First, we note that if it was the case that liquidity effects vary with local labor market conditions, we believe it is likely that liquidity constraints will tend to be more binding when local labor market conditions are poor. This will cause our estimates of how moral hazard varies with local labor market conditions to be downward biased, making it even more likely that the moral hazard cost of UI decreases with the unemployment rate. Second, we report results in Table 7 which directly address concerns about liquidity constraints. Column (1) reports our baseline specification for comparison. Columns (2) and (3) report results for subsamples where liquidity effects are likely to be less important. Column (2) focuses on the subsample of unemployed workers without a mortgage, while column (3) focuses on the subsample of unemployed workers in the 3rd and 4th quartiles of net liquid wealth. In both cases the coefficient on the interaction term is larger than in the baseline. The last two columns report results which include a full set of liquid wealth quartile dummy variables interacted with a combination of occupation fixed effects, industry fixed effects, unemployment duration, and the UI benefit level. The results consistently support the interpretation that the moral hazard cost of UI decreases with the unemployment rate.[40]

Alternative Specifications and Controls. Finally, we report additional results in

[39] Of course, the duration elasticity could vary with unobservable characteristics, though we cannot test this directly. To the extent that the distribution of these unobservable characteristics varies with local unemployment, then our estimates will include the effect of unobserved compositional changes in the sample of individuals experiencing unemployment spells.

[40] To save space, we do not report the interactions between the UI benefit level and the wealth quartile dummies in column (5), but the coefficents are very similar to Chetty (2008), implying that including local unemployment rate and its iteraction with UI benefit does not alter inference on the importance of liquidity effects.

Table 8 which vary the specification and the set of controls. In column (2), we include region-specific linear time trends and show that our result gets stronger. Column (3) includes a full set of region fixed effects interacted with year fixed effects. Identification in this specification is coming from only from variation in benefits within region-year cells. In column (4), we include state-specific linear time trends. Our main results are fairly robust to these alternative specifications. Finally, columns (5) drops the control variables; the coefficient on the interaction terms fall in magnitude by 33% and is no longer statistically significant at conventional levels ($p = 0.162$).

3.3 Calibrating the Welfare Implications

Our empirical findings suggest that moral hazard decreases with the unemployment rate. To see what this finding implies for optimal policy, we now calibrate the optimal UI level implied by our model, following the spirit of the "sufficient statistic" approach to welfare analysis. To review, this method requires using the reduced form empirical estimates as inputs into the optimal UI formula.

Our search model implies the following structural relationship for the duration elasticity:

$$\varepsilon_{D,b} = \frac{U'(b)}{U'(w_R - \tau)}\theta(w_R)ub + \frac{U'(b)}{\psi''(e)}(\delta(e))^2(1-u)b$$

One can think of $\varepsilon_{D,b} = h(u)$, where $h()$ is non-linear. In order to exploit our empirical estimates, we assume that $h()$ be locally approximated by a linear function of u. A first-order Taylor series expansion of $h(u)$ around $u = \overline{u}$ yields:

$$\varepsilon_{D,b}(u) = \varepsilon_{D,b}(\overline{u}) + \frac{d\varepsilon_{D,b}(\overline{u})}{du} \times (u - \overline{u})$$

This can also be derived directly from our reduced-form estimating equation (3.19):

$$\log h = \alpha + \beta_1 \log(b) + \beta_2 \log(b) \times (u - \overline{u}) + e \tag{3.21}$$

With this specification,

$$\varepsilon_{D,b}(u) = \frac{d\log h}{d\log(b)} = \beta_1 + \beta_2 \times (u - \overline{u})$$

Thus, $\beta_1 = \varepsilon_{D,b}(\overline{u})$ and $\beta_2 = \frac{d\varepsilon_{D,b}(\overline{u})}{du}$. Our empirical results imply that $\widehat{\beta}_1 = -0.741$ and $\widehat{\beta}_2 = .142$. To analyze the welfare implications, we will assume that the budget effect can be ignored.[41] This requires assuming that $w_R - \tau$ does not vary with u. In practice, whether the budget effect is likely to bind is related to whether a change in unemployment is temporary or permanent. If the change in unemployment is transitory, it seems safe to assume that the government wouldn't alter financing arrangements. On the other hand, moral hazard varies with unemployment regardless of whether or the change in unemployment is temporary or permanent.

Recall, the consumption smoothing benefit of UI

$$\frac{U'(b) - U'(w_R - \tau)}{U'(w_R - \tau)}$$

Assuming preferences are given by $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$, the consumption smoothing benefit is given by

$$\frac{b^{-\gamma}}{(w_R - \tau)^{-\gamma}} - 1$$

Thus, substituting this into (3.17), we get:

$$\left(\frac{w_R - \tau}{b}\right)^{\gamma} = 1 + \beta_1 + \beta_2 \times (u - \overline{u})$$

To be consistent with the calibrations above, we maintain the same parameter values.[42] Note that at these parameter values, $w_R - \tau \approx 400$. Plugging in the parameter values and solving for b yields

$$b^* = \frac{400}{(1 + \varepsilon_{D,b})^{1/1.75}}$$

[41]Incorporating the budget effect increases the complexity of the model and makes a tractable solution less easy to obtain. In ongoing work, we are working to incorporate this effect.

[42]In our data, $\overline{u} = 6.7\%$.

where $\varepsilon_{D,b} = 0.741 - 0.142 \times (u - 6.6\%)$ At $u = 6.7\%$, $b^* = 291$ implying an optimal replacement rate of 72.8%. At an unemployment rate of 8.4% (roughly one standard deviation above the mean unemployment rate), $b^* = 317$, implying a replacement rate of 79.2%. Thus, we see that variation in the unemployment rate can substantially affect replacement rates. Table 9 presents the optimal benefit level and replacement rate, for a range of unemployment rates. The basic lesson to emerge from the table is that plausible variation in the unemployment rate generates wide variation in the optimal level of UI. To give a sense of the quantitative importance of this variation, the magnitude is roughly equivalent to a one unit change in the coefficient of relative risk aversion in the model (e.g., from $\gamma = 2$ to $\gamma = 3$).

3.4 Conclusion

In this paper, we have considered a standard search model and have shown that it implies a relationship between the moral hazard cost of UI and the level of unemployment in the local labor market. This relationship is theoretically ambiguous and depends on the relative strengths of two behavioral channels: the search channel and reservation wage channel. This motivated our empirical strategy which estimated how the elasticity of unemployment duration with respect to the UI benefit level varies with the unemployment rate.

Our empirical findings indicate that moral hazard is lower when unemployment is high, consistent with the speculation of Krueger and Meyer (2002) who claimed that there is likely less of an efficiency loss from reduced search effort by the unemployed when local labor market conditions are poor. We have also shown how one can use the empirical relationship between the duration elasticity and the unemployment rate to calibrate a simple optimal UI formula.

We view the concept that the moral hazard cost of social policies may vary with local labor market conditions as possibly very general. It is plausible that the disincentive effects of other government policies may also be lower in times of high unemployment. For example, if the labor supply response to tax changes is lower during recessions, the deadweight loss of income taxation could vary with aggregate labor market conditions.

While we focused on the UI benefit level as the policy parameter, in practice, the potential benefit duration is extended during times of high unemployment. In ongoing work, we are studying theoretically how government should optimally set the potential benefit duration. This will naturally depend on the responsiveness of UI durations to changes in the potential duration parameter. We hope that this analysis will hopefully shed light on the federal supplemental benefits programs in the U.S. and other developed countries.

3.5 Appendix A: Proofs

Proof of Proposition 3.

Start by differentiating the optimal condition for search with respect to b

$$\psi''(e)\frac{\partial e}{\partial b} = \frac{\lambda''(e)}{r+s}\frac{\partial e}{\partial b}\varphi(w_R) + \frac{\lambda'(e)}{r+s}\frac{\partial \varphi(w_R)}{\partial w_R}\frac{\partial w_R}{\partial b}$$

Note that $\lambda'' < 0$, $\psi'' > 0$, and $\frac{\partial \varphi(w_R)}{\partial w_R} < 0$ so that $sign(\frac{\partial e}{\partial b}) \neq sign(\frac{\partial w_R}{\partial b})$. Next, totally differentiating the reservation wage equation with respect to b yields

$$\begin{aligned}
U'(w_R - \tau)\frac{\partial w_R}{\partial b} &= U'(b) + \frac{\partial e}{\partial b}\frac{\lambda(e)}{\lambda'(e)}\left(\psi''(e) - \psi'(e)\frac{\lambda''(e)}{\lambda'(e)}\right) \\
U'(w_R - \tau)\frac{\partial w_R}{\partial b} &= U'(b) + \frac{\partial e}{\partial b}\frac{\lambda(e)}{\lambda'(e)}\left(\psi''(e) - \frac{\lambda''(e)}{r+s}\varphi(w_R)\right)
\end{aligned}$$

where the last line made use of the FOC. Lets substitute in using the the equation above:

$$\begin{aligned}
U'(w_R - \tau)\frac{\partial w_R}{\partial b} &= U'(b) + \frac{\lambda(e)}{\lambda'(e)}\frac{\lambda'(e)}{r+s}\frac{\partial \varphi(w_R)}{\partial w_R}\frac{\partial w_R}{\partial b} \\
U'(w_R - \tau)\frac{\partial w_R}{\partial b} &= U'(b) + \frac{\lambda(e)}{r+s}\frac{\partial \varphi(w_R)}{\partial w_R}\frac{\partial w_R}{\partial b} \\
U'(w_R - \tau)\frac{\partial w_R}{\partial b} &= U'(b) - \frac{\lambda(e)}{r+s}(1 - F(w_R))U'(w_R - \tau)\frac{\partial w_R}{\partial b} \\
U'(w_R - \tau)\frac{\partial w_R}{\partial b} &= U'(b) - \frac{p(e)}{r+s}U'(w_R - \tau)\frac{\partial w_R}{\partial b} \\
U'(w_R - \tau)\frac{\partial w_R}{\partial b}(1 + \frac{p(e)}{r+s}) &= U'(b) \\
\frac{\partial w_R}{\partial b} &\approx \frac{U'(b)}{U'(w_R - \tau)}\frac{r+s}{r+s+p(e)} \\
\frac{\partial w_R}{\partial b} &= \frac{U'(b)}{U'(w_R - \tau)}u > 0
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\partial e}{\partial b}\left(\psi''(e) - \frac{\lambda''(e)}{r+s}\varphi(w_R)\right) &= \frac{\lambda'(e)}{r+s+p(e)}\frac{\partial \varphi(w_R)}{\partial w_R}\frac{U'(b)}{U'(w_R-\tau)} \\
\frac{\partial e}{\partial b}\left(\psi''(e) - \frac{\lambda''(e)}{r+s}\varphi(w_R)\right) &= -\frac{\lambda'(e)}{\lambda(e)}\frac{p(e)}{r+s+p(e)}U'(b) \\
\frac{\partial e}{\partial b}\left(\psi''(e) - \frac{\lambda''(e)}{s}\varphi(w_R)\right) &\approx -\frac{\lambda'(e)}{\lambda(e)}(1-u)U'(b) \\
\frac{\partial e}{\partial b} &= -\frac{\delta(e)(1-u)U'(b)}{\psi''(e) - \frac{\lambda''(e)}{s}\varphi(w_R)}
\end{aligned}$$

Proof of Theorem 4.

The first-order condition for the optimal benefit level is:

$$\frac{\partial w_R}{\partial b} + \frac{\partial w_R}{\partial \tau}\frac{\partial \tau}{\partial b} - \frac{\partial \tau}{\partial b} = 0$$

For simplicity, we assume that $\partial w_R/\partial \tau = \partial w_R/\partial b$ and hence $\partial D/\partial \tau = \partial D/\partial b$.[43] This implies that

$$\frac{\partial w_R}{\partial b} = \frac{\frac{\partial \tau}{\partial b}}{1+\frac{\partial \tau}{\partial b}}$$

Also, using the planner's budget constraint, we can write

$$\frac{\partial \tau}{\partial b} = \frac{D(1+\varepsilon_{D,b})}{\frac{1}{r+s} - D\varepsilon_{D,b}}$$

Thus, at an optimum:

$$\frac{\partial w_R}{\partial b} = \frac{D}{\frac{1}{r+s}+D}(1+\varepsilon_{D,b})$$

Note that for r small, this reduces to

$$\frac{\partial w_R}{\partial b} = u(1+\varepsilon_{D,b}) \tag{3.22}$$

If the left-hand side is larger than the right-hand side, a marginal increase in benefits raises the worker's after-tax reservation wage and so is welfare-improving. In

[43] This is true if individuals have CARA preferences.

other words, current benefit levels are too low. Note that this test doesn't quantify *how much* benefits should increase or decrease. This depends on how the left-hand side and the right-hand side respond to a change in benefits.

Substituting equation (3.14) into equation (3.22), we get

$$U'(b) = U'(w_R - \tau)(1 + \varepsilon_{D,b})$$

Intuitively, $U'(b)$ represents the marginal benefit of raising consumption while unemployed by \$1. The benefit increase means that taxes must be raised by more than \$1 due to the behavioral response which reduces consumption in the employed state. Rearranging this expression delivers equation (3.17).

Table 1

Calibration A: Moral Hazard and the Unemployment Rate

λ	0.1	0.09	0.08	0.07	0.06	0.05
u	4.4%	4.9%	5.6%	6.7%	8.3%	11.0%
$\varepsilon^{w_R}_{D,b}$	0.17	0.18	0.19	0.21	0.24	0.29
$\varepsilon^{e}_{D,b}$	1.42	1.49	1.58	1.70	1.82	1.90
$\varepsilon_{D,b}$	1.59	1.67	1.78	1.91	2.07	2.19

Notes:

$$\varepsilon^{w_R}_{D,b} = \theta(w_R)\frac{\partial w_R}{\partial b}b,\ \varepsilon^{b}_{D,b} = -\delta(e)\frac{\partial e}{\partial b}b,\ \varepsilon_{D,b} = \varepsilon^{w_R}_{D,b} + \varepsilon^{e}_{D,b}$$

The model is calibrated under the following assumptions:

1. Wages are log-normally distributed with mean=300 and standard deviation=240
2 $\lambda(e) = \bar{\lambda} + \lambda e,\quad \bar{\lambda} = 0.02$
3. $s = 0.003868$
4. $r = b/\mathrm{E}[w] = 0.5$
5. $\mathrm{U(c)} = \mathrm{c}^{1-\gamma}/(1-\gamma)$, $\gamma = 1.75$
6. $\psi(\mathrm{e}) = \varphi\, \mathrm{e}^{1+\kappa}/(1+\kappa)$, $\varphi = 0.06$, $\kappa = 0.2$

Table 2

Calibration B: Moral Hazard and the Unemployment Rate

λ	0.16	0.14	0.12	0.1	0.08	0.06
u	5.1%	6.3%	7.9%	9.4%	10.4%	10.7%
$\varepsilon_{D,b}^{w_R}$	0.33	0.40	0.48	0.57	0.62	0.64
$\varepsilon_{D,b}^{e}$	1.79	1.72	1.38	0.76	0.24	0.04
$\varepsilon_{D,b}$	2.12	2.11	1.86	1.33	0.86	0.68

Notes:

$$\varepsilon_{D,b}^{w_R} = \theta(w_R)\frac{\partial w_R}{\partial b}b,\ \varepsilon_{D,b}^{b} = -\delta(e)\frac{\partial e}{\partial b}b,\ \varepsilon_{D,b} = \varepsilon_{D,b}^{w_R} + \varepsilon_{D,b}^{e}$$

The model is calibrated under the following assumptions:

1. Wages are log-normally distributed with mean=300 and standard deviation=240
2 $\lambda(e) = \overline{\lambda} + \lambda e, \quad \overline{\lambda} = 0.1$
3. $s = 0.003868$
4. $r = b/\mathrm{E}[w] = 0.5$
5. $\mathrm{U(c)} = \mathrm{c}^{1-\gamma}/(1-\gamma),\ \gamma = 1.75$
6. $\psi(\mathrm{e}) = \varphi\,\mathrm{e}^{1+\kappa}/(1+\kappa),\ \varphi = 0.045,\ \kappa = 0.18$

Table 3
Descriptive Statistics

	Full Sample		State Unemp. Rate < Median		State Unemp. Rate ≥ Median	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Annual Income ($000's)	20.925	13.570	20.769	12.863	21.012	13.952
Age	37.165	11.066	36.699	11.113	37.426	11.034
Years of Education	12.171	2.877	12.151	2.820	12.183	2.909
Marital Dummy	0.616	0.486	0.610	0.488	0.619	0.486
Weekly Benefit Amount ($'s)	163.33	26.80	163.98	25.71	162.96	27.39
Replacement Rate	0.491	0.082	0.492	0.080	0.490	0.084
Unemployment Duration (weeks)	18.510	14.351	17.158	13.757	19.267	14.620
Number of Spells	4307		1545		2762	

Notes: Data are individual-level unemployment spells from 1985-2000 SIPP. Final sample of unemployment spells is described in main text.

Table 4
How does Moral Hazard vary with the
State Unemployment Rate?

		(1)	(2)
log(Average UI WBA)	**(A)**	-0.651	-0.741
		(0.318)	(0.340)
		[0.041]	[0.029]
log(Average UI WBA) × State Unemployment Rate	**(B)**		0.142
			(0.068)
			[0.038]
State Unemployment Rate		0.008	0.009
		(0.017)	(0.016)
		[0.655]	[0.598]
log(Average UI WBA) × Unemployment Duration		0.004	0.003
		(0.009)	(0.009)
		[0.674]	[0.707]
Age		-0.017	-0.017
		(0.002)	(0.002)
		[0.000]	[0.000]
Marital Dummy		0.208	0.208
		(0.040)	(0.040)
		[0.000]	[0.000]
Years of Education		0.004	0.004
		(0.006)	(0.006)
		[0.489]	[0.499]
Number of Spells		4307	4307
Post-estimation: **(A)** + σ × **(B)**			-0.502
			(0.326)
			[0.124]
Post-estimation: **(A)** - σ × **(B)**			-0.980
			(0.388)
			[0.012]

Notes: All columns report semiparametric (Cox proportional) hazard model results from estimating equation (19). Data are individual-level unemployment spells from 1985-2000 SIPP. Final sample of unemployment spells is described in the main text. Dependent variable is always the log of the individual unemployment duration (in weeks). All specifications include state, year, industry and occupation fixed effects, 10-knot linear spline in log annual wage income, controls for national unemployment rate and national unemployment rate interacted with the log of Average UI WBA and a control for being on the seam between interviews to adjust for the "seam effect." The Average UI WBA is the average weekly benefit amount paid to individuals claiming unemployment insurance. All columns estimate nonparametric baseline hazards stratified by quartile of net liquid wealth. The final two rows reports linear combinations of parameter estimates to produce the duration elasticity when the state unemployment rate is one standard deviation above/below average. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parentheses and p-values are in brackets.

Table 5
Alternative Measures of Interaction Term

	Hazard Model Results (A)	(A) × (B)	(B)	Post-estimation (A) + σ × (B)	(A) - σ × (B)
(1) **(A)** log(Average UI WBA) ×	-0.741	0.142	0.009	-0.502	-0.980
(B) State Unemployment Rate	(0.340)	(0.068)	(0.016)	(0.326)	(0.388)
	[0.029]	[0.038]	[0.598]	[0.124]	[0.012]
(2) **(A)** log(Average UI WBA) ×	-1.200	0.898	0.000	-0.301	
(B) **1**{State Unemployment Rate ≥ Median}	(0.378)	(0.262)	(0.038)	(0.310)	
	[0.002]	[0.001]	[0.996]	[0.331]	
(3) **(A)** log(Statutory Maximum UI WBA) ×	-0.269	0.120	0.004	-0.067	-0.471
(B) State Unemployment Rate	(0.314)	(0.053)	(0.018)	(0.337)	(0.316)
	[0.392]	[0.024]	[0.815]	[0.842]	[0.136]
(4) **(A)** log(Average UI WBA) ×	-0.625	0.079	0.025	-0.525	-0.725
(B) State Unemployment Rate (Finding)	(0.313)	(0.108)	(0.032)	(0.280)	(0.392)
	[0.046]	[0.462]	[0.435]	[0.061]	[0.065]
(5) **(A)** log(Average UI WBA) ×	-0.694	0.170	-0.004	-0.424	-0.964
(B) State Unemployment Rate (Separation)	(0.326)	(0.070)	(0.020)	(0.316)	(0.372)
	[0.034]	[0.016]	[0.829]	[0.179]	[0.010]
(6) **(A)** log(Average UI WBA) ×		0.209	0.026	-0.516	-1.034
(B) State Unemployment Rate (Finding)		(0.115)	(0.031)	(0.314)	(0.423)
	-0.775	[0.068]	[0.407]	[0.101]	[0.015]
(A) log(Average UI WBA) ×	(0.344)	0.243	-0.003	-0.389	-1.160
(B) State Unemployment Rate (Separation)	[0.024]	(0.081)	(0.019)	(0.328)	(0.403)
		[0.003]	[0.890]	[0.236]	[0.004]
Number of Spells	4307				

Notes: All rows report semiparametric (Cox proportional) hazard model results from estimating equation (19); each column reports separate parameter estimate. Data are individual-level unemployment spells from 1985-2000 SIPP. See Table 4 for more details on the baseline specification. The median unemployment rate across all states in sample is calculated separately each year. The Average UI WBA is the average weekly benefit amount paid to individuals claiming unemployment insurance. The State Unemployment Rate variables in rows (4) and (5) isolate variation in the unemployment driven by variation in the job finding rate and job separation rate, respectively. These variables are constructed following the method in Shimer (2007); see main text for details. The final two columns report linear combinations of the parameters. The standard deviation in the unemployment rate (σ) is 0.0168. In row (2) we set σ=1.0 becasue the interaction term includes a dummy variable rather than a continuous measure. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parentheses and p-values are in brackets.

Table 6
How Much Do Demographics Explain Why Moral Hazard Varies with the State Unemployment Rate?

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Average UI WBA)	**(A)**	-0.741	-0.719	-0.742	-0.718	-0.628	-0.618	-0.577
		(0.340)	(0.337)	(0.339)	(0.334)	(0.347)	(0.359)	(0.349)
		[0.029]	[0.033]	[0.029]	[0.032]	[0.070]	[0.085]	[0.098]
log(Average UI WBA) × State Unemployment Rate	**(B)**	0.142	0.141	0.142	0.140	0.143	0.136	0.138
		(0.068)	(0.068)	(0.068)	(0.067)	(0.070)	(0.069)	(0.068)
		[0.038]	[0.040]	[0.037]	[0.037]	[0.042]	[0.048]	[0.043]
State Unemployment Rate		0.009	0.008	0.009	0.008	0.008	0.009	0.008
		(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
		[0.598]	[0.610]	[0.598]	[0.611]	[0.605]	[0.596]	[0.606]
log(Average UI WBA) × Age			0.007					0.008
			(0.008)					(0.010)
			[0.398]					[0.428]
log(Average UI WBA) × Marital Dummy				0.020				-0.046
				(0.180)				(0.210)
				[0.912]				[0.827]
log(Average UI WBA) × Years of Education					0.049			0.051
					(0.025)			(0.031)
					[0.052]			[0.099]
Number of Spells		4307	4307	4307	4307	4307	4307	4307
log(Average UI WBA) × Occupation FEs		N	N	N	N	Y	N	Y
log(Average UI WBA) × Industry FEs		N	N	N	N	N	Y	Y
Post-estimation: **(A)** + σ × **(B)**		-0.502	-0.483	-0.504	-0.482	-0.388	-0.389	-0.345
		(0.326)	(0.327)	(0.324)	(0.320)	(0.329)	(0.336)	(0.322)
		[0.124]	[0.139]	[0.120]	[0.132]	[0.238]	[0.246]	[0.285]
Post-estimation: **(A)** - σ × **(B)**		-0.980	-0.955	-0.980	-0.953	-0.867	-0.846	-0.809
		(0.388)	(0.383)	(0.389)	(0.382)	(0.400)	(0.414)	(0.407)
		[0.012]	[0.013]	[0.012]	[0.013]	[0.030]	[0.041]	[0.047]

Notes: All columns report semiparametric (Cox proportional) hazard model results from estimating equation (19). Data are individual-level unemployment spells from 1985-2000 SIPP. See Table 4 for more details on the baseline specification. The final two rows reports linear combinations of parameter estimates to produce the duration elasticity when the state unemployment rate is one standard deviation above/below average. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parentheses and p-values are in brackets.

Table 7
Moral Hazard and Liquidity

		(1)	(2)	(3)	(4)	(5)
log(Average UI WBA)	**(A)**	-0.741	-0.780	-0.609	-0.664	
		(0.340)	(0.520)	(0.533)	(0.320)	
		[0.029]	[0.134]	[0.253]	[0.038]	
log(Average UI WBA) × State Unemployment Rate	**(B)**	0.142	0.419	0.164	0.158	0.165
		(0.068)	(0.112)	(0.127)	(0.070)	(0.074)
		[0.038]	[0.000]	[0.196]	[0.025]	[0.026]
State Unemployment Rate		0.009	0.011	-0.004	0.005	0.006
		(0.016)	(0.024)	(0.024)	(0.017)	(0.017)
		[0.598]	[0.636]	[0.852]	[0.771]	[0.727]
Number of Spells		4307	2355	2170	4307	4307
No mortgage only		N	Y	N	N	N
3rd and 4th liquid wealth quartiles only		N	N	Y	N	N
Occupation FEs × Liquid wealth quartile		N	N	N	Y	Y
Industry FEs × Liquid wealth quartile		N	N	N	Y	Y
Unemployment duration × Liquid wealth quartile		N	N	N	N	Y
log(Average UI WBA) × Liquid wealth quartile		N	N	N	N	Y
Post-estimation: **(A)** + σ × **(B)**		-0.502	-0.076	-0.333	-0.399	
		(0.326)	(0.551)	(0.553)	(0.318)	
		[0.124]	[0.890]	[0.547]	[0.210]	
Post-estimation: **(A)** - σ × **(B)**		-0.980	-1.483	-0.884	-0.929	
		(0.388)	(0.555)	(0.594)	(0.362)	
		[0.012]	[0.007]	[0.137]	[0.010]	

Notes: All columns report semiparametric (Cox proportional) hazard model results from estimating equation (19). Data are individual-level unemployment spells from 1985-2000 SIPP. See Table 4 for more details on the baseline specification. The final two rows reports linear combinations of parameter estimates to produce the duration elasticity when the state unemployment rate is one standard deviation above/below average. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parentheses and p-values are in brackets.

Table 8
Robustness to Alternative Specifications and Controls

		(1)	(2)	(3)	(4)	(5)
log(Average UI WBA)	**(A)**	-0.741	-1.010	-1.019	-1.078	-0.787
		(0.340)	(0.420)	(0.480)	(0.523)	(0.352)
		[0.029]	[0.016]	[0.034]	[0.039]	[0.025]
log(Average UI WBA) ×	**(B)**	0.142	0.157	0.156	0.151	0.095
State Unemployment Rate		(0.068)	(0.077)	(0.124)	(0.095)	(0.068)
		[0.038]	[0.041]	[0.207]	[0.113]	[0.162]
State Unemployment Rate		0.009	0.028	0.038	0.029	0.012
		(0.016)	(0.018)	(0.023)	(0.020)	(0.015)
		[0.598]	[0.116]	[0.104]	[0.134]	[0.408]
Number of Spells		4307	4307	4307	4307	4307
Baseline controls		Y	Y	Y	Y	N
Region-specific linear time trends		N	Y	N	N	N
Region × Year FEs		N	N	Y	N	N
State-specific linear time trends		N	N	N	Y	N
Post-estimation: **(A)** + σ × **(B)**		-0.502	-0.746	-0.757	-0.825	-0.627
		(0.326)	(0.420)	(0.472)	(0.551)	(0.299)
		[0.124]	[0.076]	[0.109]	[0.134]	[0.036]
Post-estimation: **(A)** - σ × **(B)**		-0.980	-1.274	-1.281	-1.331	-0.947
		(0.388)	(0.458)	(0.568)	(0.542)	(0.430)
		[0.012]	[0.005]	[0.024]	[0.014]	[0.028]

Notes: All columns report semiparametric (Cox proportional) hazard model results from estimating equation (19). Data are individual-level unemployment spells from 1985-2000 SIPP. See Table 4 for more details on the baseline specification. The final two rows reports linear combinations of parameter estimates to produce the duration elasticity when the state unemployment rate is one standard deviation above/below average. Standard errors, adjusted to allow for an arbitrary variance-covariance matrix for each metropolitan area over time, are in parentheses and p-values are in brackets.

Table 9
Model Calibrations: Optimal UI and the Unemployment Rate

u	3.3%	5.0%	6.7%	8.4%	10.1%
$\varepsilon_{D,b}$	1.218	0.979	0.741	0.503	0.264
b*	$254	$271	$291	$317	$350
r*	63.4%	67.7%	72.8%	79.2%	87.5%

Notes: All columns report optimal UI benefit levels at various levels of local unemployment. Subsequent rows report elasticity of unemployment duration with respect to UI benefit level, the optimal UI benefit level (b*) and the optimal replacement rate (r*). The optimal replacement rate is computed by dividing UI benefit level by the average wage. See Section 4 for more details on the computations.

Table A1
Calibration C: Moral Hazard and the Unemployment Rate

λ	0.095	0.09	0.085	0.08	0.075	0.07
u	4.6%	5.1%	5.8%	6.7%	7.8%	9.4%
$\varepsilon_{D,b}^{w_R}$	0.51	0.56	0.62	0.71	0.84	1.01
$\varepsilon_{D,b}^{e}$	4.47	4.76	5.07	5.37	5.57	5.48
$\varepsilon_{D,b}$	4.97	5.31	5.69	6.08	6.41	6.48

Notes:

$$\varepsilon_{D,b}^{w_R} = \theta(w_R)\frac{\partial w_R}{\partial b}b,\ \varepsilon_{D,b}^{b} = -\delta(e)\frac{\partial e}{\partial b}b,\ \varepsilon_{D,b} = \varepsilon_{D,b}^{w_R} + \varepsilon_{D,b}^{e}$$

The model is calibrated under the following assumptions:

1. Wages distributed as $F(w) = 1\text{-}(w_0/w)\text{^}(1/\sigma)$, where $w_0 = 340$ and $\sigma = 0.14$

2 $\lambda(e) = \bar{\lambda} + \lambda e,\quad \bar{\lambda} = 0.02$

3. $s = 0.003868$

4. $r = b/\mathrm{E}[w] = 0.5$

5. $\mathrm{U(c)} = \mathrm{c}^{1-\gamma}/(1\text{-}\gamma)$, $\gamma = 1.75$

6. $\psi(\mathrm{e}) = \varphi\,\mathrm{e}^{1+\kappa}/(1+\kappa)$, $\varphi = 0.035$, $\kappa = 0.1$

Table A2

Calibration D: Moral Hazard and the Unemployment Rate

λ	0.75	0.7	0.65	0.6	0.55	0.5
u	7.5%	8.4%	9.0%	9.4%	9.6%	9.7%
$\varepsilon_{D,b}^{w_R}$	1.05	1.17	1.26	1.32	1.34	1.35
$\varepsilon_{D,b}^{e}$	2.99	2.05	1.17	0.56	0.23	0.08
$\varepsilon_{D,b}$	4.04	3.23	2.43	1.87	1.57	1.44

Notes:

$$\varepsilon_{D,b}^{w_R} = \theta(w_R)\frac{\partial w_R}{\partial b} b,\ \varepsilon_{D,b}^{b} = -\delta(e)\frac{\partial e}{\partial b} b,\ \varepsilon_{D,b} = \varepsilon_{D,b}^{w_R} + \varepsilon_{D,b}^{e}$$

The model is calibrated under the following assumptions:

1. Wages distributed as $F(w) = 1\text{-}(w_0/w)\text{\^{}}(1/\sigma)$, where $w_0 = 340$ and $\sigma = 0.14$
2 $\lambda(e) = \bar{\lambda} + \lambda e,\ \ \bar{\lambda} = 0.333$
3. $s = 0.003868$
4. $r = b/\mathrm{E}[w] = 0.5$
5. $\mathrm{U(c)} = \mathrm{c}^{1-\gamma}/(1-\gamma)$, $\gamma = 1.75$
6. $\psi(\mathrm{e}) = \varphi\,\mathrm{e}^{1+\kappa}/(1+\kappa)$, $\varphi = 0.035$, $\kappa = 0.1$

Figure 1: Survival Curves Under High Unemployment

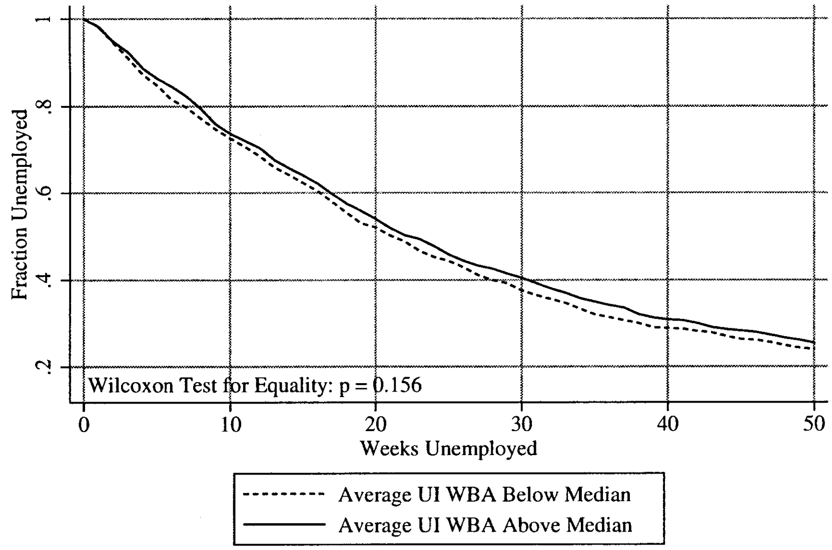
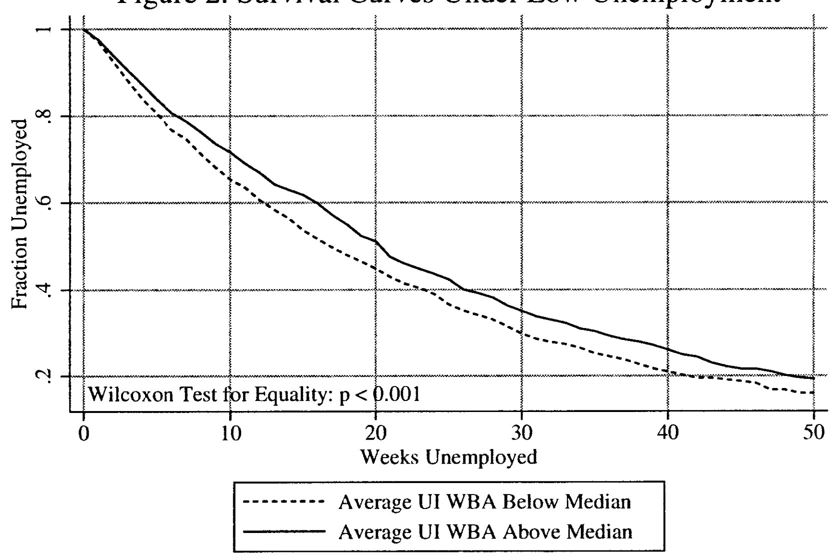


Figure 2: Survival Curves Under Low Unemployment



Notes: Data are individual-level unemployment spells from 1985-2000 SIPP. Each figure plots (Kaplan-Meier) survival curves for two groups of individuals based on whether or not Average UI Weekly Benefit Amount (WBA) in individual's state is above or below the median. The survival curves are adjusted following Chetty (2008), which parametrically adjusts for "seam effect" by fitting a Cox proportional hazard model with a seam dummy and then recovering the baseline hazard.

Figure A1: Survival Curves Under High Unemployment

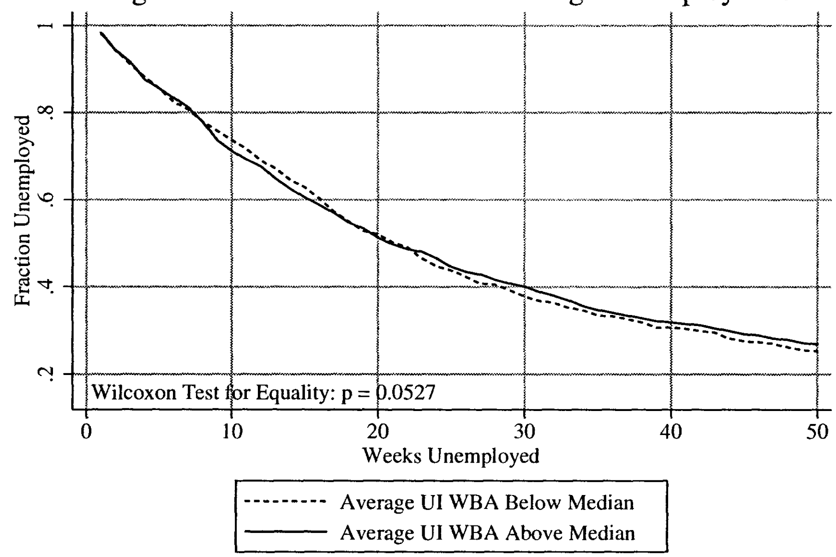
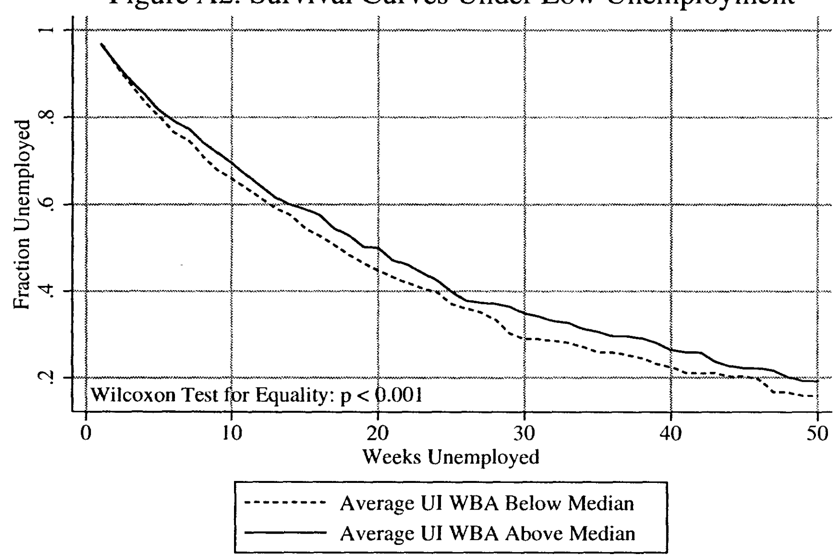


Figure A2: Survival Curves Under Low Unemployment



Notes: Data are individual-level unemployment spells from 1985-2000 SIPP, with the sample limited to unemployed workers with above-median liquid wealth. Each figure plots (Kaplan-Meier) survival curves for two groups of individuals based on whether or not Average UI Weekly Benefit Amount (WBA) in individual's state is above or below the median. The survival curves are adjusted following Chetty (2008), which parametrically adjusts for "seam effect" by fitting a Cox proportional hazard model with a seam dummy and then recovering the baseline hazard.