2019

Essays on macroeconomics with microeconomic heterogeneity

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

ESSAYS ON MACROECONOMICS WITH MICROECONOMIC HETEROGENEITY

by

WEIZE CHEN

B.A., Sun Yat-Sen University, 2010
M.A., Renmin University of China, 2013

Submitted in partial fulfillment of

requirements for the degree of

Doctor of Philosophy

2019

Approved by

First Reader

Adam Guren, PhD
Assistant Professor of Economics

Second Reader

Robert G. King, PhD
Professor of Economics

Third Reader

Marianne Baxter, PhD
Professor of Economics

Acknowledgments

I would like to thank my advisors Marianne Baxter, Adam Guren, and Robert King for their encouragement, advice, and guidence over the past years. Their influence not only shapes this dissertation but also enhances my understanding on economics and alters my way of thinking.

I am grateful to a number of faculty both at BU and from other institutions who provide valuable discussions and feedback on my dissertation. They include Martin Beraja, Ryan Chahrour, Christophe Chamley, Yanbin Chen, Stefania Garetto, Simon Gilchrist, Tarek Hassan, Zhen Huo, Eirk Hurst, Pascual Restrepo, Stephen Terry, Mark Watson, and Ivan Werning.

I also benefited from my peers and friends for many inspiring and interesting conversations: Alessandro Casini, Xiaoliang Chen, Jon Lecznar, Arthur Smith, Ei Yang, and Zhiteng Zeng.

Finally, I would like to thank my family and in particular Wendy Xu for her love and support through good times and bad times.

ESSAYS ON MACROECONOMICS WITH

MICROECONOMIC HETEROGENEITY

WEIZE CHEN

Boston University, Graduate School of Arts and Sciences, 2019

Major Professors: Adam Guren, Assistant Professor of Economics
Robert G. King, Professor of Economics

ABSTRACT

This dissertation consists of three essays on macroeconomics with microeconomic heterogeneity.

In Chapter 1, I empirically investigate the extent to which regional economic activity responds to fiscal shocks. Exploiting state-level variation of military procurement, I apply an instrumental variable local projection method extended to the panel data context to estimate the dynamic causal effects of a military spending shock. These estimates, which are referred to as "regional impulse responses," indicate three main empirical findings. First, regional output displays a large and lengthy response over a decade to a regional military spending shock, despite the fact that military spending has returned to a normal level after five years. Second, regional population gradually grows over the decade after the shock. Third, the response of construction to military spending is proportionately much larger than that of total output and also represents an important share of overall output responses. This evidence suggests that labor reallocation across regions can be very important for the impact of fiscal policy.

Chapter 2 quantitatively analyzes the regional and aggregate implications of the empirical findings in the previous chapter. I first study a simple model of regional reallocation to build intuition. Then I develop a multi-region New Keynesian model with labor migration and housing construction and calibrate a U.S. economy with 51 regions. The model reveals that labor reallocation amplifies regional output through a boom in construction spending and amplifies aggregate output through a positive "covariance effect" arising from net directed migration towards booming regions where population and regional output per resident are rising simultaneously. To circumvent high dimensionality, I propose a new method to tractably solve spatial dynamic stochastic general equilibrium (DSGE) models. Using this method, I quantitatively find that in response to a national military buildup that affects regions differentially in a manner consistent with U.S. expenditure, labor reallocation amplifies the aggregate output effect of government spending by 30 percent relative to a model without it.

Chapter 3 proposes a new method to study the macroeconomic impact of microeconomic shocks. I show in what conditions and how to represent a disaggregate stochastic dynamic model into a recursive aggregate system by approximation order. This method provides a sufficient-statistic characterization of "macro state variables or shocks" in terms of heterogeneous micro shocks and structures. The first- and second-order macro shocks can be shaped by the average and the dispersion of the micro counterparts weighted by their micro impact intensities. I apply this method in several applications to illustrate the importance of micro heterogeneity and nonlinearity in macroeconomics. First, I provide a first-order decomposition of the aggregate consumption function when consumers have different marginal propensities to consume. A new redistribution channel—the asset position adjustment channel—is identified and the magnitude of this channel relies on the variability of capital invest-

ment and the persistence of shocks. Second, I show that permanent income inequality alters aggregate demand responses mainly at the second-order, instead of the first-order. Non-homothetic preferences and the dispersion of permanent income income jointly determine the impart on aggregate consumption response.

Contents

List of Tables

List of Figures

Chapter 1

Regional Impulse Responses to Fiscal Shocks: An Empirical Investigation

1.1 Introduction

The economic consequence of fiscal policy is one of the central questions in macroeconomics. The effect of government spending on output is often summarized by a multiplier, which represents the percentage increase in output when government spending increases by one percent of GDP. This multiplier can be estimated by aggregate time series data but there is a wide range of this estimate in the literature (Ramey and Shapiro, 1988; Blanchard and Perotti, 2002; Burnside, Eichenbaum, and Fisher, 2004; Barro and Redlick, 2011; Ilzetzki, Mendoza, and Vegh, 2013).[1] The effect of government spending varies according to different macroeconomic policies, in particular monetary policy. In the past decade, a wave of empirical research (Nakamura and Steinsson, 2014; Shoag, 2016; Suarez Serrato and Wingender, 2016) exploits the variation of cross-sectional fiscal spending to estimate the "local fiscal multiplier", which measures the effect of an increase in spending in one region of a monetary union. An advantage of this cross-sectional approach is that it yields the potential for much greater variation in policy across space than over time and more plausibly exogenous variation. However, most of the studies on regional fiscal multiplier focus

[1] In this literature, some studies exploit military spending associated with wars, which are plausibly unrelated to macroeconomic conditions. Others apply the structural VAR approach to identify the multiplier.

on impact multiplier and try to translate the impact regional multiplier into the aggregate multiplier. It is still unclear how fiscal policies affect local economic activity over time and how to map regional dynamic effects to the aggregate level.

In this chapter, I empirically investigate how local economic activity responds to regional fiscal stimuli over various horizons. I exploit state-level variation of military procurement following Nakamura and Steinsson (2014) (hereafter, NS) to estimate the dynamic causal effects of a military spending shock. The econometric method is an instrumental variable local projection method as in Stock and Watson (2018) extended to the panel data context.[2] The estimates are referred to as "regional impulse responses" and are natural dynamic counterparts of NS's open economy relative multipliers.[3] As in NS, my identification takes advantage of heterogeneous state-level exposure of national military spending. For instance, when national military spending rises, state military spending increases systematically more in California than in Illinois. The regional impulse responses are inferred from differential responses of output in California relative to Illinois when military spending rises nationally.

I have three main empirical findings. First, regional output displays a large and lengthy response—it peaks at 4 years and only mean reverts after 10 years—to a regional military spending shock, despite the fact that military spending has essentially returned to a normal level after 5 years. Second, regional population gradually grows over the decade after the shock. This finding is consistent with the result in Blanchard and Katz (1992) that regional demand shocks lead to a persistent effect on regional employment, mainly due to labor reallocation across regions. Third, responses of

[2]The local projection method is originally proposed by Jorda (2005). Auerbach and Gorodnichenko (2013) and Ramey and Zubariy (2018) apply the local projection method with instruments to estimate aggregate fiscal multipliers in OECD countries and the U.S. Stock and Watson (2018) summarize some technical details when using the local projection method with external instruments.

[3]By using an instrumental variable regression for panel data, NS estimate the effect that an increase in government spending in one region of the union economy relative to another has on relative output and employment. They refer to this as the "open economy relative multiplier."

construction to military spending are proportionately much larger than those of total output and also represent an important share of overall output responses. This finding suggests that the construction sector plays an important role in shaping the aggregate response to asymmetric regional shocks when migration happens, consistent with the findings in Howard (2018). The estimated regional impulse responses to fiscal stimuli are robust across a battery of specifications, including alternative specifications by adding lag controls, alternative instruments following a Bartik approach, and placebo tests to falsify whether the baseline instruments are valid.

The rest of this chapter proceeds as follows. Section 1.2 provides some econometric foundations for my empirical investigation, which includes an illustration of local projection method with instruments for panel data, the baseline specification, and the identifying assumption. Section 1.3 displays the baseline estimates of regional impulse responses. Section 1.4 shows some robustness of regional impulse responses.

1.2 Econometric Foundations

This section first discusses how the local projection method works with instruments for panel data. The baseline specification is then provided as well as the identifying assumption and the validity of instruments.

1.2.1 Local Projection with Instruments for Panel Data

Denote $\{Y_{jt}\}$ as a stationary macroeconomic variable in region j at time t, $\{x_{jt}\}$ as the region-specific structural shocks of interest, and $\{\xi_t\}$ as economy-wide common shocks. By definition, the structural shocks $\{x_{jt}\}$ are unforecastable and uncorrelated with other region-specific structural shocks $\{v_{jt}\}$. The Slutsky-Frisch paradigm represents the path of observed variables as arising from the entire history of structural shocks. Therefore, an observed macroeconomic variable Y_{jt} can be described by a structural moving average representation in terms of the structural shocks:

$$Y_{j,t+h} = Y_j + \sum_{s=0}^{S} (\beta_s x_{j,t+h-s} + \vartheta_s v_{j,t+h-s} + \phi_s \xi_{t+h-s}), \tag{1.1}$$

where Y_j denotes a region-specific constant term, $\{x_{jt}\}$ are i.i.d. structural shocks of interest with zero mean, S is the number of lags, which can be finite or infinite, and h represents the horizon. The coefficients $\{\beta_h\}$ capture the causal effects of those structural shocks of interest over time, which are referred to as "regional impulse responses."

Let X_{jt} be an observed regressor related to the structural shock of interest x_{jt}. As pointed out by Stock and Watson (2018), the scale of x_{jt} over X_{jt} is indeterminate. This scale ambiguity is resolved by adopting, without loss of generality, a normalization for the scale of x_{jt} such that a unit increase in x_{jt} increases X_{jt} by one unit. Therefore, to estimate the dynamic causal effects $\{\beta_h\}$, a local projection regression can be specified as follows,

$$Y_{j,t+h} = \alpha_j^h + \gamma_t^h + \beta_h X_{jt} + u_{j,t+h}^h, \tag{1.2}$$

where α_j^h and γ_t^h are region and time fixed effects that absorb region-specific terms (e.g., Y_j) and economy-wide common shocks $\{\xi_{t+h-s}\}_{s\geq 0}$, and the error term u_{jt+h}^h is a linear combination of region-specific shocks $\{x_{j,t+h-s}\}_{s\neq h}$ and $\{v_{j,t+h-s}\}_{s\geq 0}$.

In general, X_{jt} is likely to be endogenous and correlated with $\{x_{j,t-s}\}_{s\geq 1}$ and $\{v_{j,t-s}\}_{s\geq 0}$. But with a suitable instrument, β_h can be consistently estimated by IV regression. Let Z_{jt} be a vector of instrumental variables. In line with Stock and Watson (2018), these instruments can be used to estimate the dynamic causal effects if they satisfy

$$
\begin{array}{rll}
\text{(i)} & \mathbb{E}(x_{jt}Z_{jt}^{'}) = \lambda^{'} \neq 0 & \text{(relevance)} \\
\text{(ii)} & \mathbb{E}(v_{jt}Z_{jt}^{'}) = 0 & \text{(contemporaneous exogeneity)} \\
\text{(iii)} & \mathbb{E}(x_{j,t-h}Z_{jt}^{'}) = 0, \forall h > 0 & \text{(lag exogeneity)} \\
 & \mathbb{E}(v_{j,t-h}Z_{jt}^{'}) = 0, \forall h > 0 & \\
\text{(iv)} & \mathbb{E}(x_{j,t+h}Z_{jt}^{'}) = 0, \forall h > 0 & \text{(lead exogeneity)} \\
 & \mathbb{E}(v_{j,t+h}Z_{jt}^{'}) = 0, \forall h > 0 & .
\end{array}
$$

Conditions (i) and (ii) are conventional IV relevance and exogeneity conditions. Condition (iii) and (iv) arise because of the dynamics. The key idea of lead/lag exogeneity conditions is that $Y_{j,t+h}$ generally depends on the entire history of structural shocks, so if Z_{jt} is to identify the effect of x_{jt} alone, it must be uncorrelated with all shocks at all leads and lags.

1.2.2 The Baseline Specification

Following NS, my empirical work employs state-level time series in military procurement, principally from the DD-350 military procurement forms of the U.S. Department of Defense for fiscal years 1966–2006. In addition, the time series data for calendar years used in this analysis includes state-level GDP constructed by the U.S. Bureau of Economic Analysis (BEA), state-level employment by the Bureau of Labor Statistics (BLS), and population by state from the Census Bureau. More details on the data set are provided in Appendix A.

I apply the local projection method discussed above to estimate the regional impulse responses to a military spending shock over various horizons h. The baseline empirical specification is[4]

$$\frac{y_{j,t+h} - y_{j,t-2}}{y_{j,t-2}} = \alpha_j^h + \gamma_t^h + \beta_h \cdot \frac{g_{jt} - g_{j,t-2}}{y_{j,t-2}} + u_{j,t+h}^h \tag{1.3}$$

where y_{jt} and g_{jt} are real output and real military spending in state j in year t respectively, and α_j^h and γ_t^h stand for state and year fixed effects. All variables in the regression are measured in per resident terms and are deflated by the national consumer price index (CPI) for the United States. The inclusion of state fixed effects absorbs state-specific time trends in output and military procurement. The year fixed effects control for aggregate shocks and aggregate monetary and tax policies. As the dependent variable in equation (1.3) is specified in difference, the coefficients of interest $\{\beta_h\}$ are *cumulated* impulse responses, which are also the same as the impulse responses for levels.[5]

NS estimate a biannual difference specification rather than a dynamic panel regression with fixed effects in annual differences, arguing that their approach is more parsimonious and mitigates measurement error in the timing of the procurement variable.[6] My baseline regressions adopt this biannual difference specification. Section 1.3 shows that the empirical results are robust to using an annual difference specification.

The dependent variable of regression (1.3) is the growth rate of regional real output relative to national price levels, which combines the impacts on regional inflation and real output relative to regional prices. In order to get rid of the regional inflation

[4]Following Hall (2009) and Barro and Redlick (2011), I construct the government spending variable by the difference of per-capita government spending relative to per-capita output in last period. This approach helps to control for heteroscedasticity especially across states in the case of panel data.

[5]Appendix B.1 shows the equivalence between cumulated impulse responses for differences and impulse responses for level.

[6]As discussed above, the procurement data are recorded for the federal government's fiscal year. Since 1976, this has been from October 1 to September 30. Prior to 1976, it was from July 1 to June 30.

impact, I adopt the state-level inflation measures in NS (2014) as a regional price deflator.[7]

Alternative Dependent Variables. To gauge the strength of labor reallocation over time, I also estimate the regional impulse responses of employment and population using an analogous approach. For employment, the regression is analogous to equation (1.3), except that the dependent variable is the growth rate of regional employment per resident, $(n_{j,t+h} - n_{j,t-2})/n_{j,t-2}$, where n_{jt} is the ratio of employment over state population. For population, the dependent variable is the growth rate of regional population, $(\mu_{j,t+h} - \mu_{j,t-2})/\mu_{j,t-2}$, where μ_{jt} is the population share in state j in year t.

1.2.3 Identifying Assumption and Validity of Instruments

Endogeneity is a key challenge to identifying the causal effects of government purchases, since military spending is notoriously political. In the presence of reverse causality between regional output and military spending, OLS estimates may be biased and inconsistent. But as NS stress, regional military spending displays systematic heterogeneous sensitivity to national military spending. These heterogeneous sensitivities across regions can be used to identify the effects of government-spending shocks. Their identifying assumption is that the United States does not embark on a military buildup because states that receive a disproportionate amount of military spending are doing poorly relative to other states.

Following this identifying assumption, I instrument for state military procurement

[7] Before 1995, they rely on state-level inflation series constructed by Del Negro (1998) for the period 1969-1995 using a combination of Bureau of Labor Statistics (BLS) regional inflation data and cost of living estimates from the American Chamber of Commerce Realtors Association (ACCRA). After 1995, they construct state-level price indexes by multiplying a population-weighted average of cost of living indexes from the ACCRA for each region with the U.S. aggregate Consumer Price Index.

in using total national procurement interacted with a state dummy. The first-stage regression is given by

$$\frac{g_{jt} - g_{j,t-2}}{y_{j,t-2}} = \delta_j + \eta_j \cdot \frac{g_t - g_{t-2}}{y_{t-2}} + \nu_t + \varepsilon_{jt}, \tag{1.4}$$

where δ_j and ν_t represent state and year fixed effects. Weak or many instruments may be a potential concern, because a large number of instruments is used in the baseline specification—one for each state. The first stage results suggest that the baseline regressions do not suffer from bias associated with weak or many instruments.[8] The Appendix also reports results using the limited information maximum likelihood (LIML) estimator, which are larger than my baseline estimates.

Another concern is autocorrelation of the identified shocks, which can cause a violation of the lead/lag exogeneity conditions. If the identified shocks are serially correlated, this may lead to a biased estimate because the identified shocks pick up some of the past variation. To address this concern, I use auxiliary regressions akin to the first-stage ones but using leads and lags of observed regressors instead.[9] Six states (Connecticut, Mississippi, Montana, North Dakota, Utah, and Virginia) fail the lead/lag exogeneity conditions and I have re-run the baseline regressions excluding those states.[10]

[8]The performance of the first-stage regression indicates that the baseline instruments satisfy relevance conditions. The R^2 is 0.26, adjusted R^2 is 0.21, and the partial R^2 is 0.12 for the baseline specification. Given a large number of instruments used in the baseline specification—one for each state, the Cragg-Donald first-stage F statistic suggested by Stock and Yogo (2005) is roughly five for the baseline specification, where the critical value for less 30 percent bias at 5 percent significance for 51 instruments is roughly 4.

[9]The auxiliary regressions are akin to the first-stage ones but using leads and lags of observed regressors instead. As the difference is constructed at a biannual frequency, the leads and lags are specified at $h = \pm 2, 4, 6, 8$ in the auxiliary regressions. Hypothetically, if the lead/lag exogeneity conditions hold, the fraction of significant slope coefficients should be small for all auxiliary regressions.

[10]The result of auxiliary regressions indicates that for each auxiliary regression, there are 4 to 8 states (out of 51) that fail the lead/lag exogeneity tests. In particular, the violation of lead/lag exogeneity conditions mainly come from the following states: Connecticut, Mississippi, Montana,

1.3 Baseline Estimates of Regional Impulse Responses

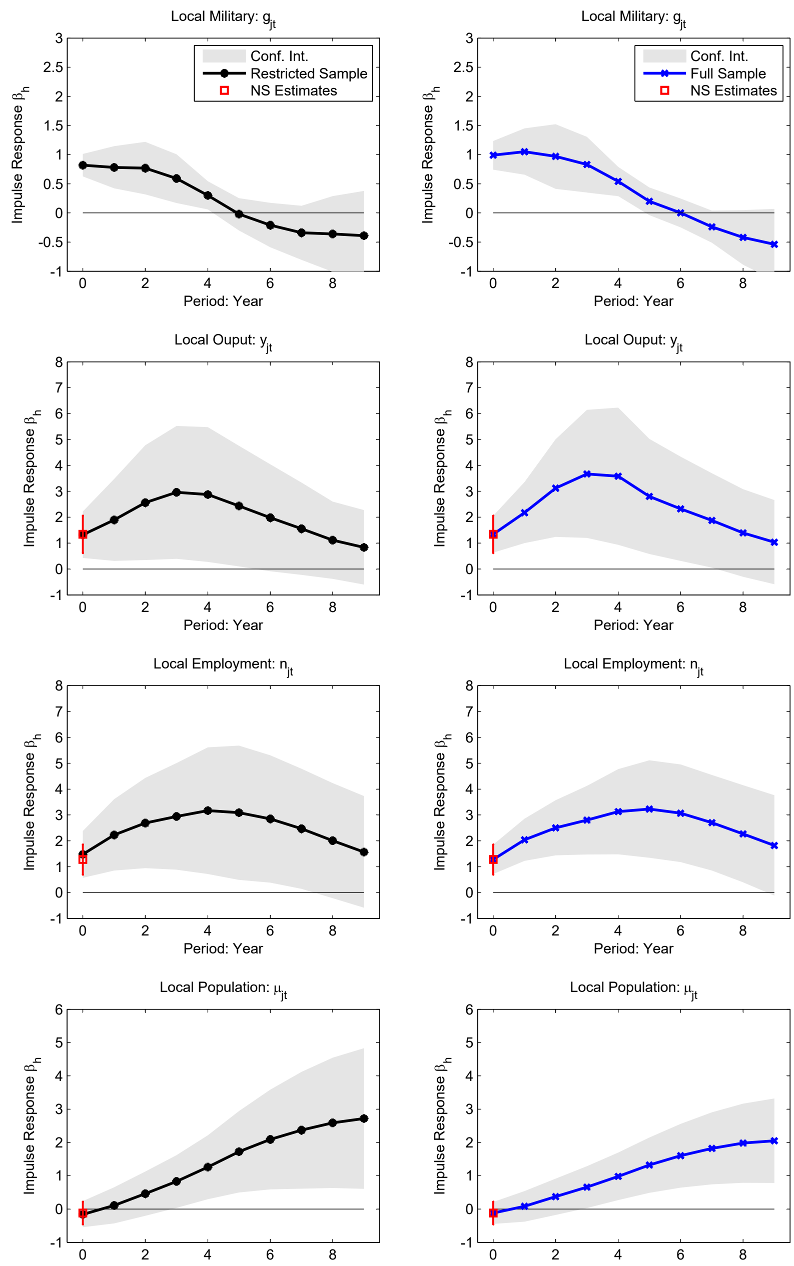
Figure 1.1 plots the baseline estimates of regional impulse responses $\{\beta_h\}$ to a military spending shock. Column (a) displays the estimates in the restricted sample, which excludes the states failing the lead/lag exogeneity tests, while Column (b) shows the estimates in the full sample.

The first row of Figure 1.1 displays the estimated regional impulse responses of military spending to a national military spending shock interacted with a state dummy. In response to a 1 percent national military spending shock relative to output, regional military purchases go up on impact and return to a normal roughly after five years. The implied AR(1) coefficient of military procurement is consistent with the NS's estimate of 0.75 at an annual frequency, indicating that the impact on regional military spending vanishes in the medium run.

1.3.1 Large and Lengthy Responses of Regional Output

The second and third rows of Figure 1.1 show that there are large responses of regional output and employment over the medium run to a military spending shock, even though the impact on regional military procurement dies out after five years. Consistent with the estimated impact multipliers in NS, state output per resident rises by 1.32 percent on impact, in response to 1 percent military spending shock relative to output. The expansionary effect on regional output persists and reaches a peak around 3 percent after four years and then gradually mean-reverts over a decade. State-level employment per resident displays a similar pattern in which regional employment goes up by 1.28 percent on impact and around 3 percent after five years. Over the medium run, the impact of government purchases on regional

North Dakota, Utah, and Virginia.



(a) Restricted Sample (b) Full Sample

Figure 1.1: Baseline Estimates of Regional Impulse Responses

Notes: The circle and cross are LP-IV estimates of regional impulse responses $\{\beta_h\}$ and the gray area is a 95 percent confidence interval. The square represents NS's estimates. Standard errors are robust and clustered by state. A shorthand for the dependent variable is stated at the top of each panel. The dependent variables are $\frac{g_{jt+h}-g_{jt-2}}{y_{jt-2}}$, $\frac{y_{jt+h}-y_{jt-2}}{y_{jt-2}}$, $\frac{n_{jt+h}-n_{jt-2}}{n_{jt-2}}$, and $\frac{\mu_{jt+h}-\mu_{jt-2}}{\mu_{jt-2}}$, where h is the horizon. The sample period is 1966-2006. Local military and output are per resident state military spending and GDP by state deflated by state-level CPI measures, employment is the number of employed divided by population. In column (a), the restricted sample excludes the states that fail the lead/lag exogeneity tests—CT, MS, MT, ND, UT, and VA; column (b) covers all states.

economic activity is even larger and lengthier than the existing literature indicated.

1.3.2 Gradual Responses of Regional Population

The fourth row of Figure 1.1 shows that estimated regional impulse responses of population are large and gradually increase over the medium run, indicating a strong regional reallocation in response to a military spending shock. As pointed out by NS, in the short run, regional military spending shocks do not cause a significant impact on regional population. However, in the medium run, regional population increases by 1 percent after five years and 2.5 percent after ten years, in response to a 1 percent military spending shock relative to output. This result is consistent with Blanchard and Katz's (1992) findings that a regional labor demand shock leads to a permanent impact on regional population and that regional reallocation is the dominant adjustment force that drives a persistent impact on regional labor markets.

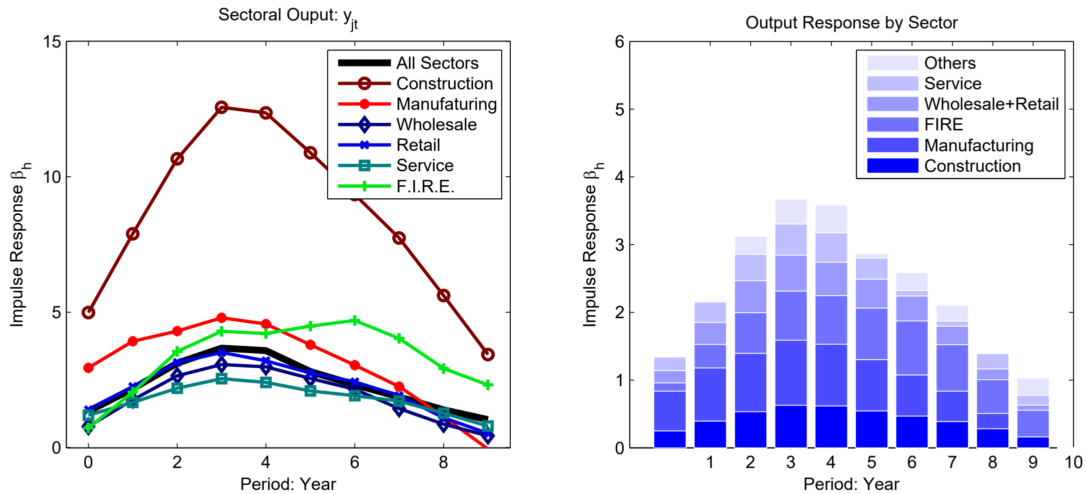
1.3.3 Particularly Large Responses in the Construction Sector

What leads to a large and lengthy response of output? To shed light on this issue, I estimate regional output impulse responses by major SIC/NAICS groupings.[11] Panel (a) in Figure 1.2 plots the regional impulse responses of sectoral output to a regional military spending shock, for construction, manufacturing, wholesale, retail, service, and FIRE (finance, insurance, rental, and estate).[12]

Panel (a) shows that responses of construction to military spending are proportionately much larger than those of total output. At the four-year horizon, the response of construction output is over 12 percent, which is roughly four times larger than the total output response. This evidence suggests that the construction sector plays

[11] Appendix C.6 also reports the estimates for sectoral employment.

[12] Other sectors—mining, agriculture, transportation and utilities, and government—do not have statistically significant responses.



(a) Responses of Sectoral Output

(b) Output Responses by Sector

Figure 1.2: Regional Impulse Responses of Output by Sector

Notes: Panel (a) shows LP-IV estimates of regional impulse responses $\{\beta_h\}$ of sectoral output and Panel (b) reports a breakdown of overall output response by sector. The dependent variable is $\frac{y^s_{jt+h} - y^s_{jt-2}}{y^s_{jt-2}}$, where h is the horizon and s denotes the sector. The sample period is 1966-2006 for all states. Sectoral output is per capita state-level GDP by sectors deflated by state-level CPI measures.

an important role in the impulse responses of regional economic activity to military spending. In addition, those sectoral estimates indicate that the large and lengthy response of output is robust across those sectors that have statistically significant responses.

Panel (b) indicates that the construction sector makes an important contribution to the overall output response. Despite the fact that the construction sector accounts for only 5 percent of total output on average, it contributes roughly 20 percent of the overall output response to a regional military spending shock. This contribution is as large as the manufacturing sector, which accounts for 20 percent of total output.[13]

[13] The output weights by sectors are 5 percent for construction, 20 percent for manufacturing, 16 percent for wholesale and retail trade, 18 percent for services, 17 percent for FIRE.

1.3.4 Summary of Empirical Results

The dynamic responses are large and lengthy, while the impact responses match those of NS. By year 4, regional output rises over 3 percent in response to 1 percent shock of local government purchases relative to output, despite the fact that government purchases have essentially returned to normal at this horizon. Further, estimated local population responses are large at the four-year horizon and continue to grow afterwards. Finally, responses of construction to government purchases are proportionately much larger than those of total output and represent an important share of overall output responses.

1.4 Robustness of Regional Impulse Responses

The estimated regional impulse responses to fiscal stimuli are robust across a battery of specifications, including (i) alternative specifications by adding controls, (ii) alternative instruments following a "Bartik" approach, and (iii) "placebo tests" to falsify whether the baseline instruments are valid. The estimated regional impulse responses with lags of dependent variables and regressors as controls are very similar to the baseline estimates. The Bartik instrument is constructed by scaling national spending for each state by the average level of spending in that state in the first five years of the sample. The estimates with Bartik instruments display similar dynamics as the ones under the baseline instruments. In the case of placebo tests, the lags of the changes in regional output do not significantly respond to the identified shocks on impact, indicating the validity of the baseline instruments. More details about robustness checks can be found in the Appendix.

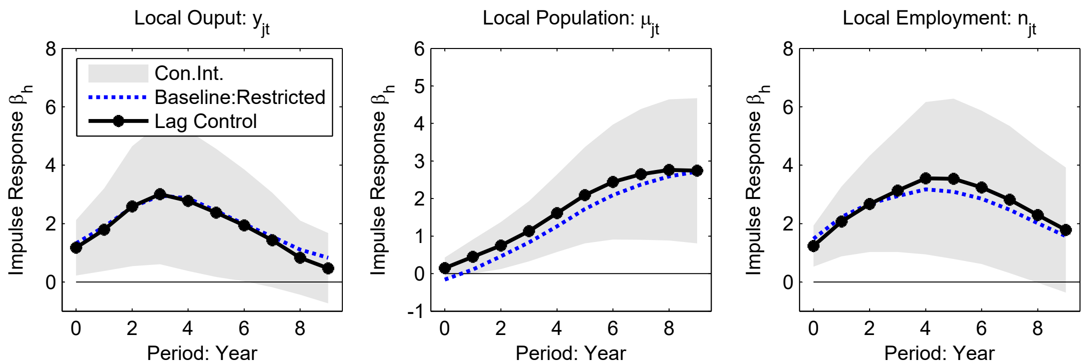
This section highlights a battery of robustness checks for estimating regional impulse responses, including (i) alternative specifications by adding controls, (ii) alternative instruments following a "Bartik" approach, (iii) "placebo tests" to falsify whether the baseline instruments are valid, and (iv) alternative specifications by using annual differences.

1.4.1 Lag Controls

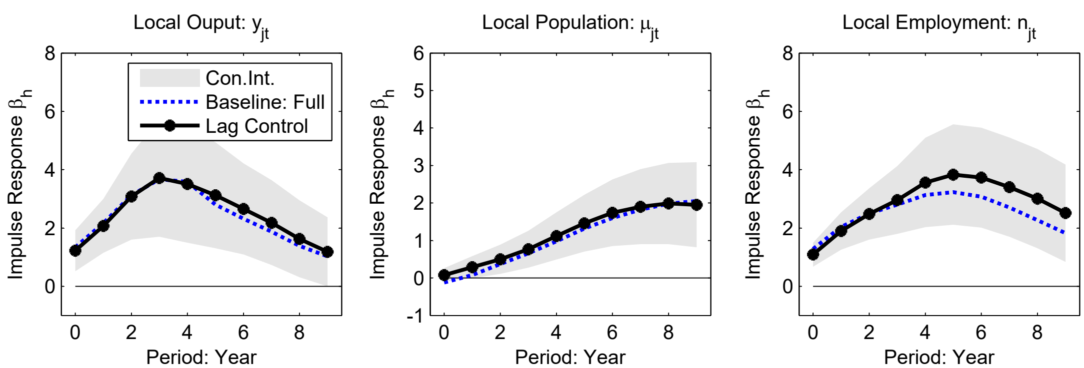
Figure 1.3 shows the estimates with the lags of regressors and dependent variables as controls. Denote $\nabla y_{jt} \equiv \frac{y_{jt}-y_{j,t-2}}{y_{j,t-2}}$ and $\nabla g_{jt} \equiv \frac{g_{jt}-g_{j,t-2}}{y_{j,t-2}}$ as the biannual changes of regional output and military spending relative to regional output. Because the shocks are constructed at a biannual frequency, the inclusion of $\nabla g_{j,t-1} = \frac{g_{j,t-1}-g_{j,t-3}}{y_{j,t-3}}$ arises colinearity with ∇g_{jt} to some extent (same as $\nabla y_{j,t-1}$ and ∇y_{jt}). Therefore, the lag controls are specified to be two-year lags instead, $\nabla g_{j,t-2}$ and $\nabla y_{j,t-2}$. The estimated regional impulse responses are very similar to the baseline estimates, indicating that the results are robust.

1.4.2 Bartik-type Instruments

This subsection employs a Bartik approach to construct instruments. The Bartik instruments are constructed by scaling national spending for each state by the average level of military spending in that state in the first five years of the sample. Given the Bartik shares, s_j, the Bartik instruments are represented by $B_{jt} = s_j \nabla g_t$, where $\nabla g_t = \frac{g_t-g_{t-2}}{y_{t-2}}$ is the biannual change of national military spending relative to national output. Row (a) in Figure 1.4 displays the estimated regional impulse responses of regional economic activity to a military spending shock using the Bartik instruments. The estimators with Bartik instruments display similar dynamics as the ones under the baseline instruments. However, the magnitude of the estimated regional impulse responses under the Bartik instruments is larger than the baseline estimates. This is probably because the Bartik instruments pick up serially correlated variation.



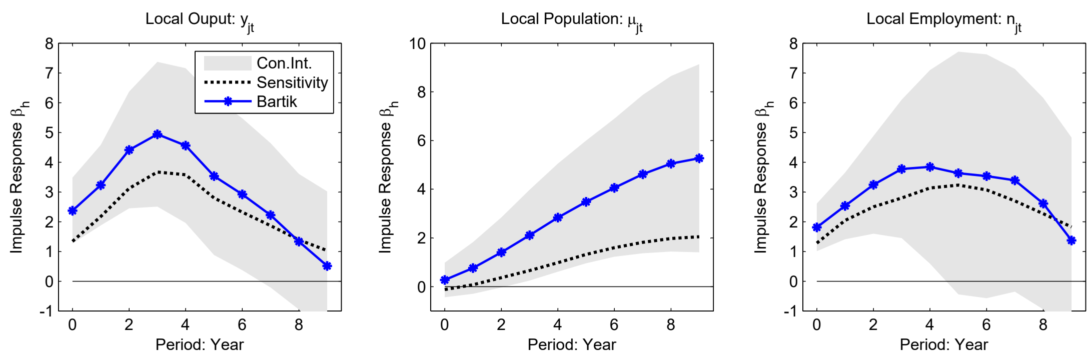
(a) Restricted Sample



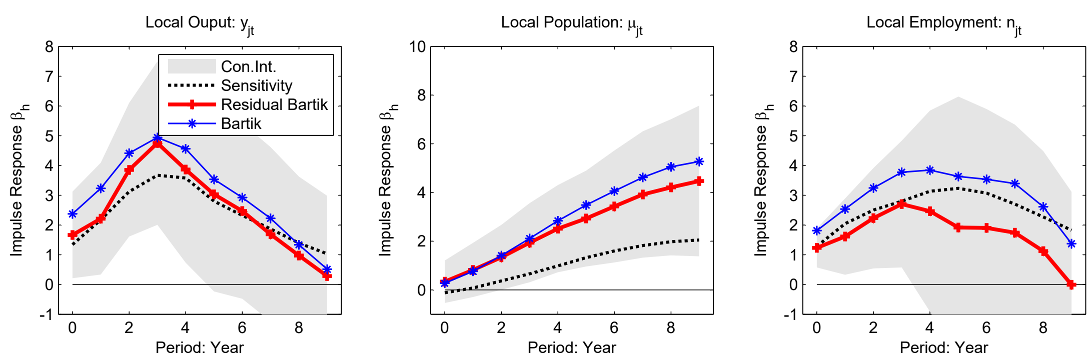
(b) Full Sample

Figure 1.3: Lags of Dependent Variables and Regressors as Controls

Notes: The circle and dot are LP-IV estimates of regional impulse responses $\{\beta_h\}$ and the gray area is a 95 percent confidence interval. Standard errors are robust and clustered by state. A shorthand for the dependent variable is stated at the top of each panel. The dependent variables are $\frac{y_{jt+h}-y_{jt-2}}{y_{jt-2}}$, $\frac{n_{jt+h}-n_{jt-2}}{n_{jt-2}}$, and $\frac{\mu_{jt+h}-\mu_{jt-2}}{\mu_{jt-2}}$, where h is the horizon. The sample period is 1966-2006. Output is per resident GDP by state deflated by state-level CPI measures, employment is the number of employed divided by population. In row (a), the restricted sample excludes CT, MS, MT, ND, UT, and VA; row (b) covers all states.



(a) Bartik Instruments



(b) Revised Bartik Instruments

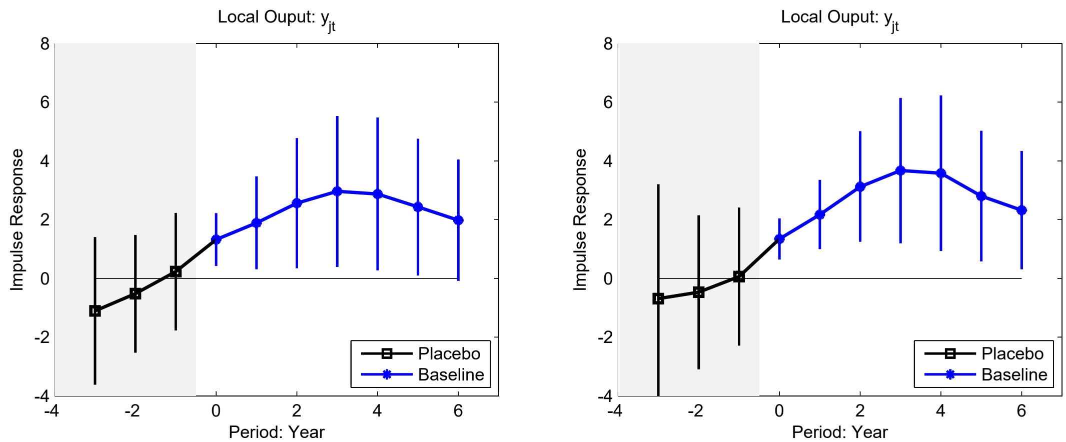
Figure 1.4: Estimated Regional Impulse Responses using Bartik Instruments

Notes: The circle and dot are LP-IV estimates of regional impulse responses $\{\beta_h\}$ with Bartik-type instruments and the gray area is a 95 percent confidence interval. Standard errors are robust and clustered by state. A shorthand for the dependent variable is stated at the top of each panel. The dependent variables are $\frac{y_{jt+h}-y_{jt-2}}{y_{jt-2}}$, $\frac{n_{jt+h}-n_{jt-2}}{n_{jt-2}}$, and $\frac{\mu_{jt+h}-\mu_{jt-2}}{\mu_{jt-2}}$, where h is the horizon. The sample period is 1966-2006 for all states. Output is per resident GDP by state deflated by state-level CPI measures, employment is the number of employed divided by population.

In order to get rid of the possible serially correlated variation in the Bartik instruments, I construct revised Bartik instruments B_{jt}^{res} as the residuals from the following regression, $B_{jt} = \delta + \lambda_1 B_{jt-2} + \lambda_2 B_{jt+2} + \epsilon_{jt}$. Therefore, the revised Bartik instruments B_{jt}^{res} are uncorrelated with the lead and lag with two periods. Row (b) in Figure 1.4 displays the estimated regional impulse responses under revised Bartik instruments. The estimated regional impulse responses under the revised Bartik instruments are closer to the baseline estimates, relative to the estimates under original Bartik instruments.

1.4.3 Placebo Tests

Placebo (or falsification) tests are usually applied to assess whether the instruments are valid or note. Hypothetically, if the baseline instruments are valid, the identified shocks $\widehat{\nabla g_{jt}}$ (which are purified by the baseline instruments) should be uncorrelated with the changes of regional output before time t. Therefore, one can run the baseline regression but with the changes of regional output before time t as dependent variables instead, i.e., $\nabla \tilde{y}_{j,t-s} = \alpha_j + \gamma_t + \eta_s \cdot \widehat{\nabla g_{jt}} + \upsilon_{jt}^s$, where $\nabla \tilde{y}_{j,t-s} = \frac{y_{j,t-1-s} - y_{j,t-3-s}}{y_{j,t-3-s}}$ denotes the biannual changes of regional output at horizon s in the past. Due to the biannual difference specification, the placebo tests start with $\frac{y_{j,t-2} - y_{j,t-4}}{y_{j,t-4}}$ i.e., $s = -1$ to -3. The black squares in Figure 1.5 denote the point estimates of $\{\eta_s\}$ and the associated vertical lines represent the 95 percent confidence intervals. The blue circles denote the point estimates of baseline estimates. This result indicates that the lags of the changes in regional output do not significantly respond to the identified shocks on impact, which is supportive evidence for the validity of the baseline instruments.



(a) Restricted Sample (b) Full Sample

Figure 1.5: Placebo Tests

Notes: The black squares denote the point estimates of $\{\eta_s\}$ and the associated vertical lines denote the 95% confidence intervals. Blue circles denote the point estimates of baseline results, $\{\beta_h\}$. The dependent variables are $\frac{y_{jt+h}-y_{jt-2}}{y_{jt-2}}$, where h is the horizon. The sample period is 1966-2006. Output is per resident GDP by state deflated by state-level CPI measures. In panel (a), the restricted sample excludes CT, MS, MT, ND, UT, and VA; Panel (b) covers all states.

1.4.4 Annual Difference Specifications

The baseline specification follows NS's biannual specification to mitigate potential measurement errors in the timing of procurement variables. This subsection discusses the robustness of annual difference specification. Using annual differences, the regression is specified as $\frac{y_{jt+h}-y_{jt-1}}{y_{jt-1}} = \alpha_j + \gamma_t + \beta_h \cdot \frac{g_{jt}-g_{jt-1}}{y_{jt-1}} + u_{jt+h}$ and the instruments are total national procurement $\frac{g_t - g_{t-1}}{y_{t-1}}$ interacted with a state dummy.

The statistical performance of the first-stage regression and auxiliary regressions can be used to gauge whether the validity conditions—relevance and lead/lag exogeneity conditions—hold for annual difference specification. The first-stage performance is fair, but not as good as the one for the biannual difference specification. The R^2 is 0.17 and the partial R^2 is 0.088. The Cragg-Donald F statistic is 3.67. Table 1.1 reports the fraction of significant slope coefficients in the auxiliary regressions for the

	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6
Fraction	4%	6%	8%	8%	8%	10%

	Lead1	Lead2	Lead3	Lead4	Lead5	Lead6
Fraction	8%	8%	8%	10%	6%	8%

Table 1.1: Fraction of significant slope coefficients: Annual Difference

Notes: The auxiliary regressions are specified by $\frac{g_{jt}-g_{jt-1}}{y_{jt-1}} = \alpha_j^g + \gamma_t^g + \eta_j^s \cdot \frac{g_{t+s}-g_{t+s-1}}{y_{t+s-1}} + u_{jt}^s$, where $s = \pm 1, 2, 3, 4, 5, 6$. The table reports the fraction of significant (at the 95% level) slope coefficient η_j^s in the auxiliary regressions. The total number of auxiliary regressions is 51, i.e., one for each state.

annual changes of national military spending. It shows that the lead/lag exogeneity conditions are likely to hold, while the strength of instrument relevance might not be as strong as the baseline ones.

Figure 1.6 shows the estimated regional impulse responses in the annual difference specification. The estimates display similar patterns to the baseline results over various horizons. The major difference is the point estimation on impact: output and employment multiplier on impact are 1.34 and 1.28 in biannual difference specification while they are 0.64 and 0.39 in annual difference specification. These differences arise because the strength of instrument relevance in annual difference specification is not as strong as the baseline instruments. As NS point out, annual difference specification might be subject to measurement error in the timing of being recorded and being carried out actually. In particular, the military procurement data is measured in federal government fiscal year while output and employment are measured in calendar year. It may cause measurement errors in the timing. But overall, the estimated regional impulse responses display similar patterns to the baseline estimates over various horizons.

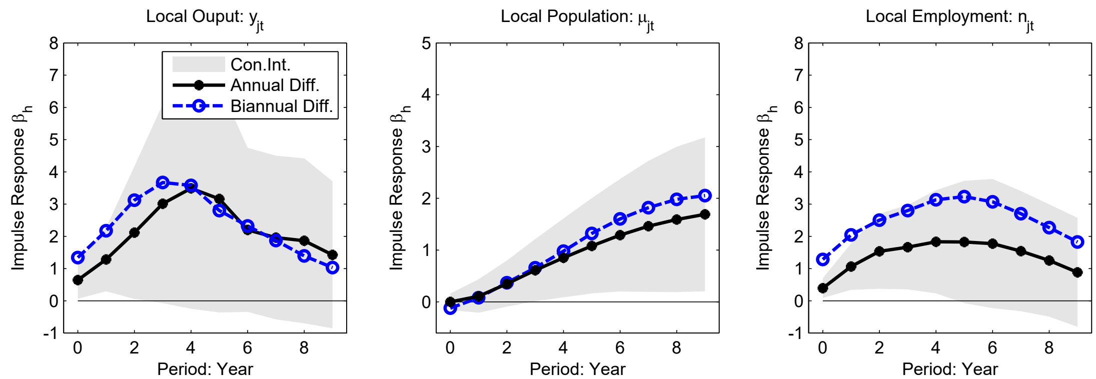


Figure 1.6: Estimated Local Multipliers in Annual Difference Specification

Notes: The circle and dot are LP-IV estimates of regional impulse responses $\{\beta_h\}$ for the annual and biannual difference specifications. The gray area is a 95 percent confidence interval for the estimates under annual difference specification. A shorthand for the dependent variable is stated at the top of each panel. The dependent variables are $\frac{y_{jt+h}-y_{jt-2}}{y_{jt-2}}$, $\frac{n_{jt+h}-n_{jt-2}}{n_{jt-2}}$, and $\frac{\mu_{jt+h}-\mu_{jt-2}}{\mu_{jt-2}}$, where h is the horizon. The sample period is 1966-2006 for all states. Output is per resident GDP by state deflated by state-level CPI measures, employment is the number of employed divided by population.

1.5 Conclusion

In this chapter, I exploit state-level variation of military procurement to estimate the dynamic causal effects of a military spending shock. The estimated regional impulse responses indicate that the regional economic consequences of government spending are large and lengthy and dynamics are very important for regional impulse responses to fiscal shocks. My empirical findings provide a new set of facts that can potentially discipline structural models to map regional impulse responses to the aggregate level. In Chapter 2, I develop a quantitative multi-region model to assess the importance of regional reallocation on the aggregate effect of fiscal policy.

Appendix

A. Data

Military Procurement. The main source for military spending data is the electronic database of DD-350 military procurement forms from the U.S. Department of Defense. They cover all prime contracts greater than $10,000 from 1966 to 1983 and greater than $25,000 to 2006. This data is for the federal government's fiscal year.[14] This data is available from the data package of Nakamura and Steinsson (2014). All the military procurement contracts are measured in units of million of current dollars. All prime contracts of military purchases are compiled by state for the federal government's fiscal year.

Gross Domestic Output (GDP). The state-level output is measured by GDP by state from the Bureau of Economic Analysis (BEA). It is annual data from 1963 and in millions of current dollars. GDP by state is measured as the factor incomes incurred in production, as is Gross Domestic Income (GDI). Although GDP by state is measured like GDI, the factor income are reconciled with GDP as the final step in the estimation process. The estimation of GDP by state is the sum of the ones over all industries. Estimates for 1963-1997 is based on the Standard Industrial Classification (SIC); while after 1997, it is based on the 2007 North American Industry Classification System (NAICS). The sectoral GDP by state is measured by major SIC/NAICS groupings, including agriculture, construction, FIRE (finance, insurance, rental, estate), government, manufacturing, mining, retail trade, services, transportation and utilities, and wholesale trade.

[14] Since 1976, fiscal year has been from October 1 to September 30. Prior to 1976, it was from July 1 to June 30.

Employment. There are two available time series of employment by state: total non-farm employees by state from the Bureau of Labor Statistics (BLS) and employment (measured as the number of jobs) by state from BEA. The total non-farm employees by state from BLS is available from 1939 (for most of states, the rest begins from 1960) to present. The BEA measure of state employment is available from 1969. It provides total employment measured by the number of jobs, non-farm employment (by industry), and wage and salary employment (by type).

Population. The population by state is available from the Census Bureau from 1900 to present. Between census years, population is estimated using a variety of administrative data sources including birth and death records, IRS data, Medicare data, and data from the Department of Defense. Since 1970, it is also available to obtain population by age group, which allows us to construct estimates of the working age population. Notably, the migration data from IRS starts from 1980s, which provides a short time series relative to the Census population data by state.[15]

GDP Deflator. The implicit price deflator is constructed by BEA with account code, A191RD3, which is an annual index and normalized by 2009 dollar. The state-level inflation measures are the ones adopted by Nakamura and Steinsson (2014) as a local price deflator. Before 1995, they rely on state-level inflation series constructed by Del Negro (1998) for the period 1969-1995 using a combination of BLS regional inflation data and cost of living estimates from the American Chamber of Commerce Realtors Association (ACCRA). After 1995, they construct state-level price indexes by multiplying a population-weighted average of cost of living indexed from the ACCRA

[15]The IRS migration data is available for the following periods: 1978-1980 (in-migration only), 1980-1981, 1983-1984, and 1984 to present for each tax year. https://www.archives.gov/research/electronic-records/reference-report/irs-data.html

for each state with the U.S. aggregate Consumer Price Index.

B Econometrics Foundations

B.1 The Method of Instrumental Variable Local Projection

This subsection first shows the derivation of the validity conditions for an instrumental variable local projection (LP-IV) method for panel data: the relevance, contemporaneous exogeneity, lead/lag exogeneity conditions. The second part shows the equivalent relationship between cumulative impulse responses for differences and impulse responses for levels.

B.1.1 Validity Conditions

Denote $\{Y_{jt}\}$ as a stationary macroeconomic variable in region j at time t, $\{x_{jt}\}$ as the region-specific structural shocks of interest. By definition, structural shocks $\{x_{jt}\}$ are unforecastable and uncorrelated with other shocks $\{\upsilon_{jt}\}$. An observed stationary variable Y_{jt} can be described by a structural moving average representation in terms of structural shocks,

$$Y_{j,t+h} = Y_j + \sum_{s=0}^{S} (\beta_s x_{j,t+h-s} + \vartheta_s \upsilon_{j,t+h-s} + \phi_s \xi_{t+h-s}) \tag{1.5}$$

where Y_j denotes a region-specific constant term, S is the number of lags, which can be finite or infinite, and h represents the horizon. The coefficients $\{\beta_h\}$ capture the causal effects of those structural shocks of interest over time, which is referred as to "regional impulse responses."

Let X_{jt} be an observed regressor related to the structural shock of interest x_{jt}. As pointed out by Stock and Watson (2018), the scale of x_{jt} over X_{jt} is indeterminate.

Without loss of generality, this scale ambiguity is resolved by adopting a unit increase in x_{jt} increases X_{jt} by one unit. Assume X_{jt} can be represented by $X_{jt} = x_{jt} + \omega_{jt}$, where ω_{jt} is the component independent to x_{jt} and could be correlated with $\{x_{j,t-s}\}_{s\geq 1}$ and $\{\upsilon_{j,t-s}\}_{s\geq 1}$ and includes measurement errors that are independent to all structural shocks.

Therefore, the above structural moving average representation can be rewritten as

$$Y_{j,t+h} = Y_j + \beta_h X_{jt} + \sum_s \phi_s \xi_{t+h-s} + \left(\sum_{s\neq h} \beta_s x_{j,t+h-s} + \sum_s \vartheta_s \upsilon_{j,t+h-s} - \beta_h \omega_{jt} \right) \quad (1.6)$$

This implies the local projection regression

$$Y_{j,t+h} = \alpha_j^h + \beta_h X_{jt} + \gamma_t^h + u_{j,t+h}^h \quad (1.7)$$

where α_j^h and γ_j^h are region and time fixed effects that absorb region-specific terms (Y_j) and economy-wide common shocks $(\sum_s \phi_s \xi_{t+h-s})$, and the error term $u_{j,t+h}^h$ is a linear combination of region-specific $\{x_{j,t+h-s}\}_{s\neq h}$ and $\{\upsilon_{j,t+h-s}\}_{s\geq 0}$ as well as measurement error terms.

In general, X_{jt} is likely to be endogenous and thus ω_{jt} could be correlated with $\{x_{j,t-s}\}_{s\geq 1}$ and $\{\upsilon_{j,t-s}\}_{s\geq 1}$. But with a suitable instrument, β_h can be consistently estimated by IV regression. Let Z_{jt} be a vector of instrument variables. The instruments are valid if they satisfy the conditions (i) relevance $\mathbb{E}(X_{jt}Z_{jt}^{'}) \neq 0$ and (ii) exogeneity $\mathbb{E}(u_{j,t+h}^h Z_{jt}^{'}) = 0$. In particular, since the error term $u_{j,t+h}^h$ is a linear combination of region-specific $\{x_{j,t+h-s}\}_{s\neq h}$ and $\{\upsilon_{j,t+h-s}\}_{s\geq 0}$, the exogeneity condition implies three types of exogeneity conditions:

$$\begin{aligned}
\mathbb{E}(v_{jt}Z'_{jt}) &= 0 && \text{(contemporaneous exogeneity)} \\
\mathbb{E}(x_{j,t-h}Z'_{jt}) &= 0, \forall h > 0 && \text{(lag exogeneity)} \\
\mathbb{E}(v_{j,t-h}Z'_{jt}) &= 0, \forall h > 0 && \\
\mathbb{E}(x_{j,t+h}Z'_{jt}) &= 0, \forall h > 0 && \text{(lead exogeneity)} \\
\mathbb{E}(v_{j,t+h}Z'_{jt}) &= 0, \forall h > 0 && .
\end{aligned}$$

Thus, the relevance condition also implies that

$$\mathbb{E}(x_{jt}Z'_{jt}) \neq 0. \text{ (relevance)}$$

B.1.2 Cumulative Impulse Responses for Differences

In many applications, the dependent variable is specified in first differences, e.g., $\Delta Y_{j,t+\tau} \equiv Y_{j,t+\tau} - Y_{j,t+\tau-1}$, but our interest is impulse responses for levels. Existing literature has pointed out the equivalence that cumulated impulse responses for differences are the same as impulse responses for levels. I now show this equivalent relationship for the instrumental variable local projection method for panel data.

The local projection regression for the dependent variable in first differences is given by

$$\Delta Y_{j,t+\tau} = \tilde{\alpha}_j^\tau + \tilde{\gamma}_t^\tau + \delta_\tau \cdot x_{jt} + \tilde{u}_{j,t+\tau}^\tau,$$

and thus the regression for the dependent variable in cumulated differences is given by

$$\sum_{\tau=0}^{h} \Delta Y_{j,t+\tau} = \sum_{\tau=0}^{h} \tilde{\alpha}_j^\tau + \sum_{\tau=0}^{h} \tilde{\gamma}_t^\tau + \sum_{\tau=0}^{h} \delta_\tau \cdot x_{jt} + \sum_{\tau=0}^{h} \tilde{u}_{j,t+\tau}^\tau. \tag{1.8}$$

Claim: The cumulated impulse responses for differences $\sum_{\tau=0}^{h} \delta_\tau$ are the same as the impulse responses for levels β_h.

Proof: According to the local projection regression for the dependent variable for levels,

$$Y_{j,t+\tau} = \alpha_j^\tau + \gamma_t^\tau + \beta_h x_{jt} + u_{j,t+\tau}^\tau,$$

the regression for the dependent variable in first differences can be represented by

$$\Delta Y_{j,t+\tau} = (\alpha_j^\tau - \alpha_j^{\tau-1}) + (\gamma_t^\tau - \gamma_t^{\tau-1}) + (\beta_\tau - \beta_{\tau-1}) \cdot x_{jt} + (u_{j,t+\tau}^\tau - u_{j,t+\tau-1}^{\tau-1}).$$

In particular, if $\tau = 0$, this yields

$$\Delta Y_{jt} = (\alpha_j^0 - \alpha_j^{-1}) + (\gamma_t^0 - \gamma_t^{-1}) + (\beta_0 - 0) \cdot x_{jt} + (u_{jt} - u_{j,t-1}^{-1}),$$

for the reason that x_{jt} has no impact on Y_{jt-1}.

Therefore, the cumulated impulse responses for differences

$$\sum_{\tau=0}^{h} \delta_\tau = \sum_{\tau=0}^{h} (\beta_\tau - \beta_{\tau-1}) = \beta_h,$$

which are equivalent to the impulse responses for level, β_h. **Q.E.D.**

B.2 Discussions on Regression Specifications

B.2.1 Variable Construction: Multipliers vs. Elasticities

Ramey (2016) suggests two alternative ways to construct regression variables for estimating government spending multipliers. One is based on Hall (2009) and Barro and Redlick (2011), while the other one accords to Gordon & Krenn (2010). The Hall-Barro-Redlick (HBR) transformation constructs variables as $\hat{w}_{t+h} \equiv (w_{t+h} -$

$w_{t-1})/y_{t-1}$, where w_t is the National Income and Product Account (NIPA) per-capita variable deflated by the GDP deflator and y_{t-1} is real GDP per capita in the last period $t-1$. The Gordon-Krenn (GK) transformation divides all NIPA per-capita variables by "potential GDP," estimated as an exponential trend. Thus, the NIPA variables are transformed to be $\hat{w}_t = w_t/y_t^\star$, where $y_t^\star$ is the estimated trend in real GDP per capita. In most of cases, both methods give similar results. Following the HBR approach, the government spending multiplier can be estimated by

$$\left(\frac{y_t - y_{t-1}}{y_{t-1}}\right) = \gamma + \beta^y \cdot \left(\frac{g_t - g_{t-1}}{y_{t-1}}\right) + \varepsilon_t \tag{1.9}$$

where β^y is the coefficient of interest, implying how much dollar of output increases as government spending rises by one dollar on average, i.e., $\beta^y = \frac{\Delta y}{\Delta g}$, which is the "multiplier."

Notably, a standard way in previous literature to define left-hand-side variable is as the log-difference of government spending, $\log(g_t/g_{t-1}) \approx \frac{g_t - g_{t-1}}{g_{t-1}}$. However, the estimated coefficient does not directly reveal the government spending multiplier because the percent changes must be converted to dollar equivalents. Most of the analyses using log-difference method obtain the spending multiplier by using an *ad hoc* conversion factor based on the sample average of y/g. More precisely, they follows a regression model,

$$\log\left(\frac{y_t}{y_{t-1}}\right) = \gamma + \delta^y \cdot \log\left(\frac{g_t}{g_{t-1}}\right) + \varepsilon_t \tag{1.10}$$

where $\log(y_t/y_{t-1}) \approx \frac{y_t - y_{t-1}}{y_{t-1}}$ and δ^y implies the "elasticity", $\frac{\Delta y/y}{\Delta g/g}$, therefore $\beta^y = \delta^y \cdot (y/g)$. According to Owyang, Ramey and Zubairy (2013), their investigations reveal that this widely-used method can lead to biases in multiplier estimates. In particular, they find that this method often generates multipliers greater than unity even when auxiliary specifications show that private spending falls when government spending increases. This bias occurs because the ratio of y/g varies greatly over the sample period they consider. In addition, this bias may be severer in the context of panel data because of regional heteroscedasticity.

The HBR transformation has several advantages relative to the log-difference transformation: (i) it is a direct way to obtain the government spending multiplier in dollar equivalents; (ii) it avoids potential upward bias due to ignoring the variation of y_t/g_t over the sample period when using *ad hoc* conversion factor; (iii) it also helps to control heteroscedasticity, in particular, across states in case of panel data.

For non-NIPA variables, e.g., employment, the standard way is to convert them as log-difference, $\frac{n_t - n_{t-1}}{n_{t-1}}$, where n_t denotes working hours or employment per capita. Thus, the regression to estimate the government spending multiplier on employment follows

$$\left(\frac{n_t - n_{t-1}}{n_{t-1}}\right) = \gamma + \beta^n \cdot \left(\frac{g_t - g_{t-1}}{y_{t-1}}\right) + \varepsilon_t$$

Therefore, β^n implies the change of employment growth rate given one percentage increase of government spending relative to output. Alternatively, one could also employ growth rate of government spending $(g_t - g_{t-1})/g_{t-1}$ as left-hand-side variable such that the multiplier can be estimated by $(n_t - n_{t-1})/n_{t-1} = \gamma + \delta^n \cdot (g_t - g_{t-1})/g_{t-1} + \varepsilon_t$. However, β^y and δ^n are not comparable for the reason that β^y measures percentage change of output as government spending rises by one percent *relative to output* while δ^n measures the one of output as government spending rises by one percent. Therefore, if β^y is estimated by using HBR transformation, then it is better to keep using HBR transformed variables to obtain the multipliers of other variables. This procedure enables to avoid similar biases due to ignoring the variation of y_t/g_t over the sample period when using *ad hoc* conversion factor.

B.2.2 Coefficient Interpretations

Given the baseline specification with the HBR transformation,

$$\frac{y_{j,t+h} - y_{j,t-2}}{y_{j,t-2}} = \alpha_j^h + \gamma_t^h + \beta_h \cdot \frac{g_{jt} - g_{j,t-2}}{y_{j,t-2}} + u_{j,t+h}^h,$$

the coefficients of interest $\{\beta_h\}$ can be interpreted as the impulse responses in the multiplier sense. In response to a dollar shock of regional government spending per resident, regional output per resident rises by β_h dollar at horizon h after the shock. Notably, the one dollar shock will also lead to a persistent increase in regional government spending per resident. Alternatively, $\{\beta_h\}$ can be also interpreted as follows. In response to a 1 percent shock of regional government spending relative to regional output, regional output per resident rises by β_h percent at horizon h relative to the pre-shock level. In the calibration, the parameters are calibrated by matching the model-based impulse responses to 1 percent regional government spending shock and the estimated impulse responses $\{\hat{\beta}_h\}$.

For alternative variables, e.g., regional employment rates, the coefficients of interest $\{\beta_h^n\}$ can be interpreted as follows. In response to 1 percent shock of regional government spending relative to output, regional employment per resident rises by β_h^n percent at horizon h relative to the pre-shock level. The interpretation is similar in the case of population, sectoral output, and sectoral employment as alternative dependent variables.

C More Discussions on Empirical Results

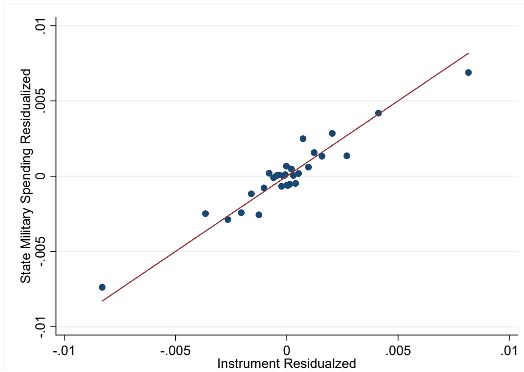
C.1 Visual Representations

This subsection provides a visual representation of the baseline specification. Figure 1.7 represents binned scatter plots for the "first stage" regression and "reduced form" regressions of the IV elasticities of per-resident state output and population over different horizons, respectively. The x variable is the residualized instrument, and y variables are the residualized regional military spending, regional output, and regional population in the biannual difference specification over various horizons. Both the x and y variables are demeaned by year and state fixed effects. These plots show that neither the first-stage nor the reduced-formed relationships are driven by outliers.

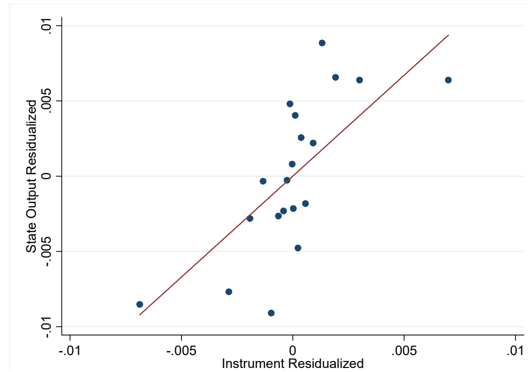
C.2 Validity of Baseline Instruments

First, the first-stage performance can be used to gauge the strength of instrument relevance. R^2 is 0.26; adjusted R^2 is 0.21; and the partial R^2 of the excluded instruments is 0.12 for the baseline specification. Given that a large number of instruments is used in the baseline specification—one for each state, the Cragg-Donald (1993) first-stage F statistic suggested by Stock and Yogo (2005) is roughly five for the baseline specification, where the critical value for less than 30 percent bias at 5 percent significance for 51 instruments is roughly 4. This evidence suggests that the baseline instruments satisfy relevance conditions.

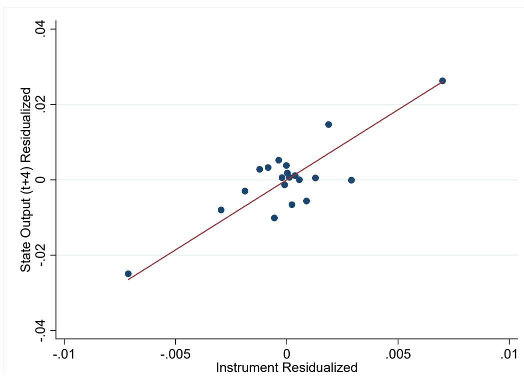
Second, the contemporaneous exogeneity condition is not directly testable. The analysis relies on the identifying assumption—national military spending does not respond because states that receive a disproportionate amount of military spending are doing poorly relative to other states. If the identifying assumption is correct, then $\mathbb{E}(v_{jt+h}Z'_{jt}) = 0, \forall h$, i.e., contemporaneous exogeneity conditions and lead/lag



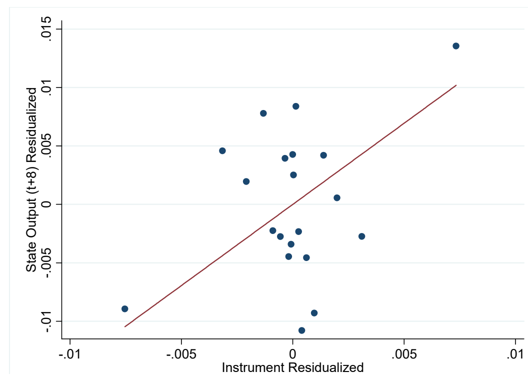
(a) First Stage



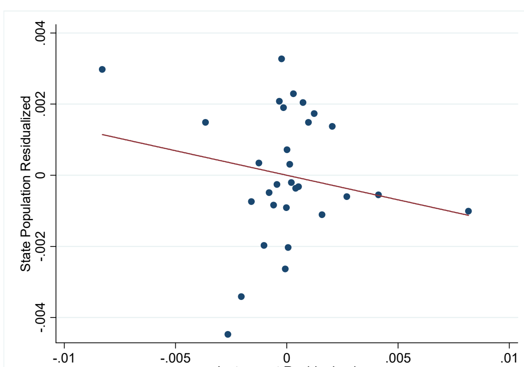
(b) State Output $\frac{y_{jt}-y_{j,t-2}}{y_{j,t-2}}$



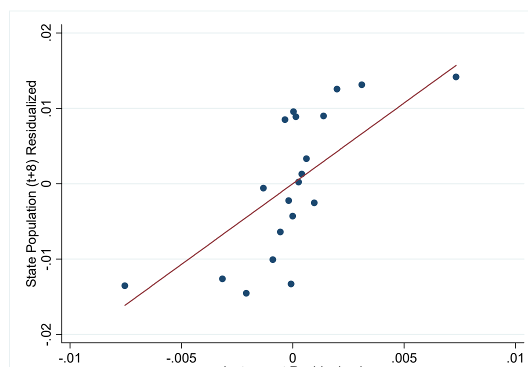
(c) State Output $\frac{y_{j,t+4}-y_{j,t-2}}{y_{j,t-2}}$



(d) State Output $\frac{y_{j,t+8}-y_{j,t-2}}{y_{j,t-2}}$



(e) State Population $\frac{\mu_{jt}-\mu_{j,t-2}}{\mu_{j,t-2}}$



(f) State Population $\frac{\mu_{j,t+8}-\mu_{j,t-2}}{\mu_{j,t-2}}$

Figure 1.7: First Stage and Reduced Form Binned Scatter Plots

Notes: The figure shows binned scatter plots of the first stage and reduced form of the IV elasticities of per-resident state output and population over different horizons. Both the x and y variables are demeaned by year and state fixed effects. The sample period is 1966-2006 for all states. Local military and output are per resident state military spending and GDP by state deflated by state-level CPI measures.

exogeneity conditions for other region-specific structural shocks $\{v_{jt}\}$ hold.

Third, I run auxiliary regressions with leads and lags of the biannual changes in national military spending interacted with a state dummy, to "test" whether lead/lag exogeneity conditions for the structural shocks of interest hold or not. Denote the biannual changes of regional military spending and of national military spending by $\nabla g_{jt} \equiv \frac{g_{jt}-g_{jt-2}}{y_{jt-2}}$ and $\nabla g_{t+h} \equiv \frac{g_{t+h}-g_{t+h-2}}{y_{t+h-2}}$, respectively. The lead/lag exogeneity conditions imply that ∇g_{jt} and $\mathbb{I}_j \nabla g_{t+h}$ are uncorrelated for any $h \neq 0$, where $\mathbb{I}_j$ denotes a state dummy. That's to say, the slope coefficients should be not significant for any horizon $h \neq 0$. Since a large number of instruments—one for each state—are employed in each auxiliary regression, the fraction of significant slope coefficients in each auxiliary regression is a concise index to show the performance. Hypothetically, if the lead/lag exogeneity conditions hold, the fraction of significant slopes should be small for all auxiliary regressions.

As the difference is constructed in a biannual frequency, the auxiliary regressions are run with leads and lags at $h = \pm 2, 4, 6, 8$. Table 1.2 reports the fraction of significant slope coefficients in the auxiliary regressions. The result indicates that for each auxiliary regressions, there are 4-8 states (out of 51) that fail the lead/lag exogeneity tests. Since the fraction is not an ordinary statistic, one cannot judge whether it is large or small in a statistical sense. However, the tests are still able to provide useful information. I find that six states fail most of the lead/lag exogeneity tests: Connecticut, Mississippi, Montana, North Dakota, Utah, and Virginia. That's to say, the leads and lags of the national military spending are correlated with the current changes of local military spending in these states. Hence, a natural way to check robustness is to re-run the baseline regressions by excluding these states.

	Lag2	Lag4	Lag6	Lag8	Lead2	Lead4	Lead6	Lead8
Fraction	16%	12%	12%	6%	16%	12%	8%	8%

Table 1.2: Fraction of Significant Slope Coefficients in Auxiliary Regressions

Notes: The auxiliary regressions are specified by $\frac{g_{jt}-g_{jt-2}}{y_{jt-2}} = \alpha_j^g + \gamma_t^g + \eta_j^s \cdot \frac{g_{t+s}-g_{t+s-2}}{y_{t+s-2}} + u_{jt}^s$, where $s = \pm 2, 4, 6, 8$. The table reports the fraction of significant slope coefficient η_j^s in the auxiliary regressions. The total number of auxiliary regressions is 51, i.e., one for each state.

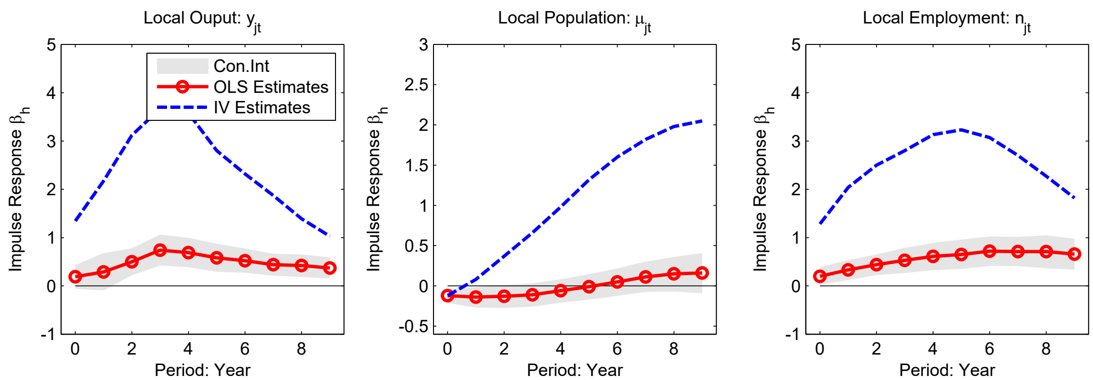


Figure 1.8: Regional Impulse Responses: IV vs. OLS Estimates

Notes: The circle and dot are point estimates of regional impulse responses $\{\beta_h\}$ in using OLS and IV respectively. A shorthand for the dependent variable is stated at the top of each panel. The dependent variables are $\frac{y_{jt+h}-y_{jt-2}}{y_{jt-2}}$, $\frac{n_{jt+h}-n_{jt-2}}{n_{jt-2}}$, and $\frac{\mu_{jt+h}-\mu_{jt-2}}{\mu_{jt-2}}$, where h is the horizon. The sample period is 1966-2006. Output is per resident GDP by states deflated by state-level CPI measures, employment is the number of employed divided by population.

C.3 Comparing with OLS Estimates

Figure 1.8 represents the estimated regional impulse responses based on a local projection regression with and without instruments. Blue dashed lines represent the baseline estimates of regional impulse responses (with the baseline instruments), while red circles stand for the estimated impulse responses in using OLS. The OLS estimates have a similar shape pattern but much smaller, reflecting the expected endogeneity bias.

C.4 The Specification Regional Total Term

In the baseline specification, state-level output and military spending are measured in per capita terms. Hence, the dependent variable $\frac{y_{jt+h}-y_{jt-2}}{y_{jt-2}}$ and the regressor

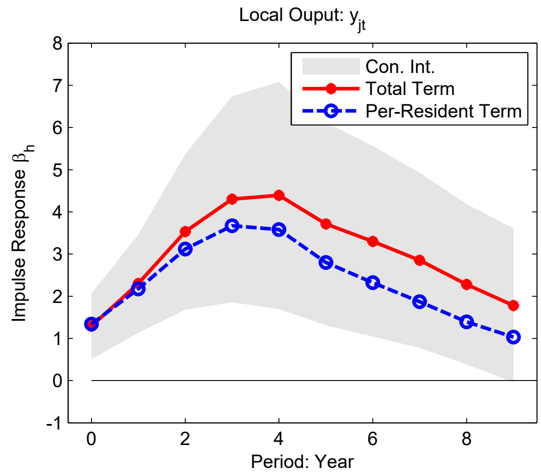


Figure 1.9: Regional Total Terms vs. Per-Resident Terms

Notes: The circle and dot are LP-IV estimates of regional impulse responses $\{\beta_h\}$ in regional per resident terms and in regional total terms respectively. The dependent variable is $\frac{y_{jt+h}-y_{jt-2}}{y_{jt-2}}$, where h is the horizon. The sample period is 1966-2006 for all states. Output is per resident GDP by state deflated by state-level CPI measures.

$\frac{g_{jt}-g_{jt-2}}{y_{jt-2}}$ are interpreted as net change rates of variable per capita. The regional impulse response β_h represents how much the net change rate of per capita output is in response to 1 percent increase of localized military spending relative to its output.

Alternatively, one can employ state-level output and military spending measured in *regional total* terms and estimate the effects on net change rate of state output in total terms. Hence, the regression specification is given by

$$\frac{\tilde{y}_{jt+h}-\tilde{y}_{jt-2}}{\tilde{y}_{jt-2}} = \alpha_j + \gamma_t + \beta_h \cdot \frac{\tilde{g}_{jt}-\tilde{g}_{jt-2}}{\tilde{y}_{jt-2}} + u_{jt+h}$$

where $\tilde{y}_{jt}$ and $\tilde{g}_{jt}$ are the regional total level of real output and military spending. I thus instrument for the biannual change rates of state military procurement $\frac{\tilde{g}_{jt}-\tilde{g}_{jt-2}}{\tilde{y}_{jt-2}}$ in using the biannual change rates of national procurement in regional total level $\frac{\tilde{g}_{t}-\tilde{g}_{t-2}}{\tilde{y}_{t-2}}$ interacted with a state dummy. Figure 1.9 reports the estimates. Since the change rate of state output in total levels contains the change rate of per capita output and of regional population. As expected, the estimated regional impulse responses in total level terms are systematically larger than the ones in per capita terms, especially after 5 years as regional populations grow substantially. The results are consistent with our conjecture, indicating significant labor migration towards booming regions.

C.5 Long-Run Average Multipliers

In the baseline specification, I apply an instrumental variable local projection method to estimate the regional impulse responses to fiscal stimuli. Usually, one can apply long differences of regional output and military spending to estimate the "long-run average multipliers" in using the following regression specification

$$\frac{y_{j,t+h} - y_{j,t-2}}{y_{j,t-2}} = \alpha_j^h + \gamma_t^h + \delta_h \cdot \frac{g_{j,t+h} - g_{j,t-2}}{y_{j,t-2}} + u_{j,t+h}^h \tag{1.11}$$

where δ_h is usually interpreted as the long-run average multiplier. Figure 1.10 reports the estimates of δ_h under the baseline instrument specification. For a better comparison, I also compute the implied long-run average multipliers based on the estimated regional impulse responses by $\delta_h^{implied} = \frac{1}{h}\sum_{s=0}^{h} \hat{\beta}_s$, where $\hat{\beta}_h$ are the point estimates of regional impulse responses in the baseline specification. The estimates for the variables in regional per resident terms are shown in the left panel and the ones in regional total terms in the right panel.

C.6 Regional Impulse Responses of Sectoral Employment

The left panel of Figure 1.11 reports the employment multipliers across sectors. The employment multipliers are similar to the output multipliers in the corresponding sectors. The local employment in construction sector also sharply responds to local military spending shocks. Other sectors including mining, agriculture, transportation

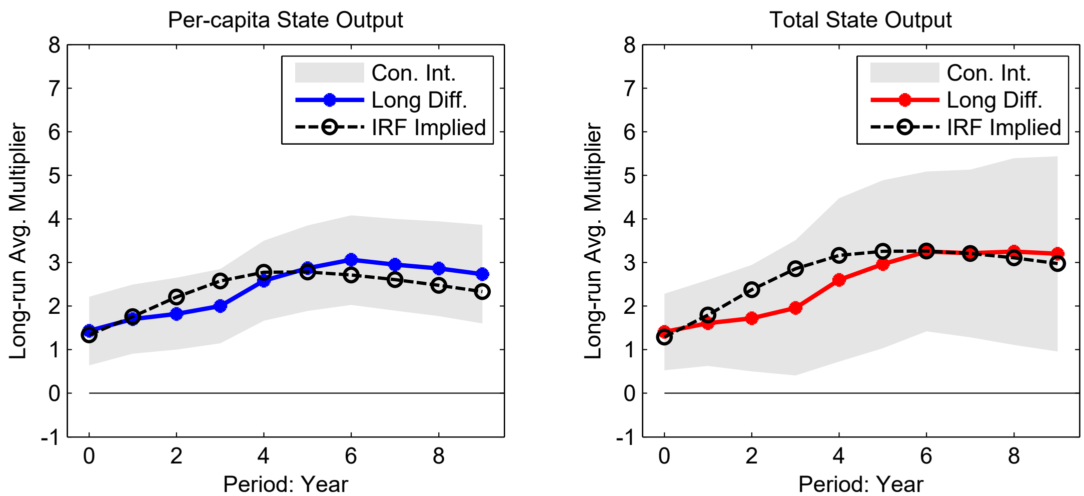
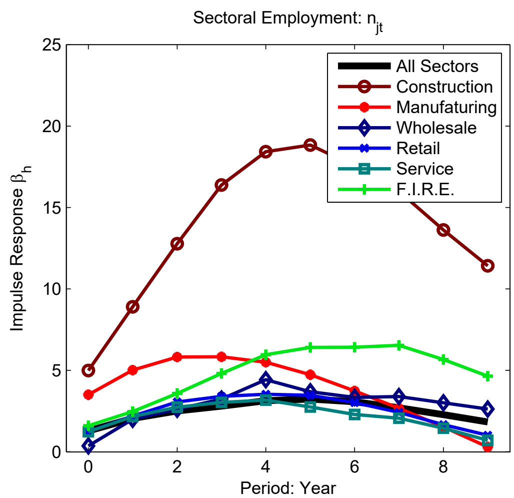


Figure 1.10: Long-run Average Multipliers

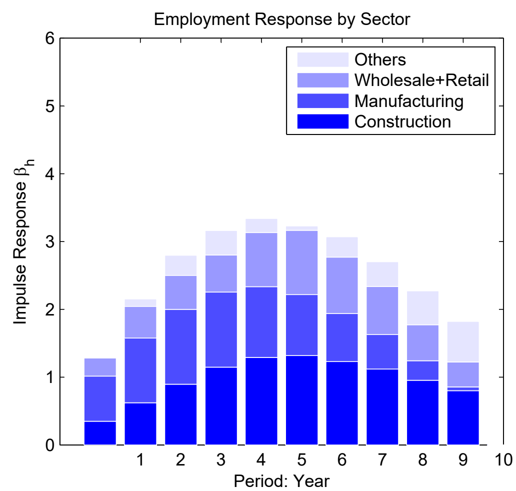
Notes: The dots are point estimates of long-run average output multipliers $\{\delta_h\}$ in regional per resident terms and in regional total terms respectively for the left and right panels, as in regression (1.11). The sample period is 1966-2006 for all states. The circles are the implied long-run average output multipliers $\delta_h^{implied} = \frac{1}{h}\sum_{s=0}^{h}\hat{\beta}_s$, where $\hat{\beta}_h$ are the LP-IV estimates of regional impulse responses in the baseline specification.

and utilities, and government do no have statistically significant responses neither. The right panel indicates that the construction sector makes an important contribution to the overall employment response.[16]

[16]The sectoral employment shares are 7 percent for construction, 19 percent for manufacturing, 9 percent for FIRE, 6 percent for wholesale trade, 21 percent for retail trade, and 31 percent for services.



(a) Responses of Sectoral Employment



(b) Employment Responses by Sector

Figure 1.11: Regional Impulse Responses of Employment by Sectors

Notes: The panel (a) shows LP-IV estiamtes of regional impulse responses $\{\beta_h\}$ of sectoral employment, which is measured by the number of employed divided by regional population. Panel (b) reports a breakdown of employment responses by sector. The dependent variable is $\frac{n^s_{jt+h} - n^s_{jt-2}}{n^s_{jt-2}}$, where h is the horizon and s denotes the sector. The sample period is 1966-2006 for all states.

Chapter 2

Quantifying Fiscal Multipliers by a World of Regional Reallocation

2.1 Introduction

The regional exposure of a macroeconomic shock or policy intervention is often substantially asymmetric across regions (i.e., regions have idiosyncratic sensitivities.) These asymmetric impacts may trigger regional reallocation—labor migration across regions—that affects both the regional and aggregate dynamics of the economy. While a literature initiated by Blanchard and Katz (1992) has emphasized that regional reallocation is a primary margin of adjustment,[1] there is essentially no quantitative modeling of the role of regional reallocation in determining aggregate economic activity.[2]

In this chapter, I study how labor reallocation across regions in response to fiscal stimuli affects *regional* and *aggregate* economic activity quantitatively. According to the previous chapter, regional output and population display large and lengthy responses to a regional fiscal stimulus, especially in the construction sector. To elu-

[1] Also see Eichengreen (1993), Borjas (2001), Glaeser and Gottlieb (2009), Kennan and Walker (2011). Labor migration is a significant force by which the U.S. labor market adjusts to regional shocks, despite a recent slowdown in regional reallocation (e.g., Kaplan and Schulhofer-Wohl, 2017).

[2] Existing literature (e.g., Nakamura and Steinsson, 2014; Farhi and Werning, 2017) mainly focuses on a monetary union model assuming that regions are small open economies where goods and capital are freely traded but population is fixed across regions. One exception is Farhi and Werning (2014), who study the role of labor migration in a static monetary union model. Another is House, Proebsting, and Tesar (2018) who introduce labor mobility into a multi-country DSGE model to quantify the benefits labor mobility in a currency union.

cidate the underlying mechanism and to quantify the aggregate impact, I develop a multi-region New Keynesian model with labor migration and housing construction and calibrate a U.S. economy with 51 regions. The model reveals that labor reallocation amplifies regional output through a boom in construction spending and amplifies aggregate output through a positive "covariance effect" arising from net directed migration towards booming regions where population and regional output per resident are rising simultaneously. To circumvent high dimensionality, I propose a new method to tractably solve spatial DSGE models. Using this method, I quantitatively find that in response to a national military buildup that affects regions differentially in a manner consistent with U.S. expenditure, labor reallocation amplifies the aggregate output effect of government spending by 30 percent relative to a model without it.

This chapter quantitatively analyzes the regional and aggregate implications of my empirical findings. However, before proceeding to a complex quantitative model, I study a simple model of regional reallocation to build intuition. This example features asymmetric regional government purchase shocks and assumes that labor is freely mobile across regions. The simple model reveals that labor migration and housing construction are each crucial to generate amplification effects on both regional and aggregate responses to fiscal stimuli. In response to a regional fiscal expansion, labor shifts to booming regions. Workers migrating to a region bring not only their labor supply but also their demand for local housing. The increase in regional housing demand leads to a more than proportional rise in construction spending, because immobile factors, such as land, are in fixed supply. Therefore, the construction response triggered by labor migration amplifies the impact of regional fiscal stimulus on *regional* output per resident.

More importantly, regional reallocation amplifies the *aggregate* output effect of asymmetric localized government spending shocks. Intuitively, in response to asymmetric regional shocks, population moves towards booming regions. These regions experience an output and construction boom which increases output per resident. The combination of population and output per resident growth occurring in the same areas generates a covariance effect that amplifies the impacts on aggregate output.

To quantitatively evaluate the aggregate impacts of labor reallocation across regions, I develop a New Keynesian monetary union model in the tradition of Benigno and Benigno (2003), Gali and Monacelli (2008), NS (2014), and Beraja, Hurst, and Ospina (2018).[3] The model consists of multiple regions in which capital and intermediate goods are freely traded. However, I relax the assumption that all regions are small open economies with fixed populations and allow individual agents to optimally decide where to move subject to migration frictions. The individual's problem is closely related to the competitive/directed labor search model of Lucas and Prescott (1974). To generate gradual labor migration as in the data, individual agents are subject to a constant "Calvo probability" of being able to relocate and idiosyncratic taste shocks for various regions. In addition, construction completion lags help to generate hump-shaped responses of construction as in the empirical work.

In general, solving a spatial DSGE model is challenging because the entire cross-sectional distribution of regional economic conditions evolves as a state variable, causing a high-dimensional state space. I propose a new method that surmounts this roadblock in order to characterize and solve the equilibrium in spatial DSGE models. Under appropriate assumptions,[4] the aggregate responses in the model un-

[3]The regional reallocation mechanism does not depend on nominal rigidity, which is shown by the simple model in Section 2.2. The purpose of introducing the New Keynesian framework is to incorporate monetary policy reaction to generate realistic aggregate fiscal multipliers.

[4]The key assumptions for aggregation include i) the underlying production, utility, and market structure are symmetric across regions and ii) the Gumbel specification of location-preference shocks. More details are discussed in Section 2.3.

der flexible perturbation can be characterized by an aggregate dynamic system with a new low-dimensional state variable capturing aggregate effects of reallocation. Applying this aggregation result, I show that the regional responses in the model can be represented by a regional dynamic system conditional on the aggregates.[5] Thus, the model-based regional impulse response is conceptually consistent with the estimated regional impulse response that includes time fixed effects soaking up all aggregate variation.

I calibrate the model to a 51-state U.S. economy. After setting some parameters to standard values, I fit the structural parameters associated with labor migration and housing construction by matching the regional impulse responses of population and construction output estimated from the data. Although I do not target the estimated responses to regional output and consumption, the model matches those impulse responses quite well. In contrast, a comparable model without labor migration does not generate lengthy hump-shaped responses of regional output, population, construction, and consumption.

Equipped with the calibrated economy, I investigate the macroeconomic implications of regional reallocation in response to fiscal stimuli. Using the systematic heterogeneity in regional sensitivity to national military spending pointed out by NS, I quantify the impacts of 1 percent national military buildup relative to output that affects state military spending differentially. Relative to a comparable model with fixed populations, regional reallocation amplifies the aggregate cumulative output multiplier by 30 percent.

The degree of amplification provided by regional reallocation depends crucially

[5]This approach shares similar advantages with Beraja, Hurst, and Ospina (2018) in that it enables us to separately solve regional and aggregate dynamics by using perturbation methods. Unlike their method, I allow for labor mobility across regions and thus aggregate variables depend on the endogenously varying population distribution across regions.

on the asymmetry and persistence of the spending. I perform two sets of quantitative experiments to demonstrate this point. First, by altering regional sensitivity to national military spending, I quantify the macro impacts in response to a national military buildup that affects regions more asymmetrically than in the benchmark case. The results show that the aggregate output effect of government purchases is greater when there are more asymmetric localized fiscal shocks across regions. This is because more asymmetric shocks trigger a larger covariance effect of regional reallocation. This dispersion dependence is in contrast to existing models with regional fixed populations, which instead predict irrelevance to the degree of asymmetry of regional shocks. Second, I investigate the role of the persistence of fiscal shocks, given the degree of asymmetry. More persistent fiscal shocks lead to a larger magnitude of labor reallocation and therefore a larger amplification effect on aggregate output.

These results indicate that analyses of fiscal policy need to consider the asymmetry and persistence of the policy in order to assess the importance of the regional reallocation channel.

The remainder of this chapter proceeds as follows. Section 2.2 studies a simple model of regional reallocation to build intuition. Section 2.3 develops a quantitative multi-regional New Keynesian model with labor migration and housing construction. Section 2.4 describes a new method to characterize and solve for equilibrium in the model. Section 2.5 calibrates parameters. Section 2.6 inspects the main mechanisms and quantifies the impact of regional reallocation on aggregate fiscal multipliers. Section 2.7 concludes.

2.2 A Simple Static Model of Regional Reallocation

Having shown the regional impulse responses to fiscal stimuli, I investigate the regional and aggregate implications of the above empirical findings. But before proceeding to a complex quantitative model, I study a simple model of regional reallocation to build intuition and illustrate the covariance effect. To that end I provide an analytical solution to equilibrium output in response to regional government spending by specifying particular functional form assumptions, as well as a general solution by using a second-order approximation for the case without functional form assumptions.

2.2.1 Basic Setup

The economy is static and consists of J regions indexed by $j \in \mathcal{J} = \{1, 2, ..., J\}$, with regional populations denoted by μ_j and total population is normalized to one, $\sum_j \mu_j = 1$. Individual households are freely mobile across regions. Lower-case variables are in per-resident terms, and upper-case variables stand for regional totals.[6] In each region, individual households have utility over consumption c_j and hours worked n_j, denoted by $u(c_j, n_j)$, where utility is concavely increasing in c_j and convexly decreasing in n_j. Each region has a competitive firm that produces regional goods by using regional labor following a production function $y_j = a_j f(n_j)$, where a_j is productivity level and f is increasing in n_j. Assume that each household demands one unit of housing inelastically. Houses are produced by using regional goods x_j and land per resident $l_j = L/\mu_j$, where L is the regional total amount of land, which is common across regions. The housing production function is denoted by $h_j = h(x_j, l_j)$, where h is increasing in x_j and l_j. Regional goods are not traded across regions and are used for regional consumption $C_j = c_j\mu_j$, construction spending $X_j = x_j\mu_j$, and government

[6] Since total population is normalized to 1, aggregate total variables equal to aggregate per capita variables.

purchases G_j, where $G_j = G + \varepsilon_j$ and ε_j are exogenous shocks. The regional goods market clears such that $y_j\mu_j = c_j\mu_j + x_j\mu_j + G_j$. The aggregate amount of government purchases is denoted by $G = \sum_j G_j$. Assume a symmetric steady state across regions such that $y_j^\star = y^\star$, $c_j^\star = c^\star$, $n_j^\star = n^\star$, $x_j^\star = x^\star$, $G_j^\star = G^\star$, $a_j^\star = a^\star$, and $\mu_j^\star = \mu^\star = 1/J$.

2.2.2 Analytical Solution

To show the regional reallocation mechanism analytically, I assume a quadratic utility function $u(c,n) = \frac{1}{2}(\bar{c} - c)^2 - \frac{\lambda_n}{2}n^2$, a linear production of regional goods $y_j = an_j$, and a Cobb-Douglas housing production, $h_j = (\frac{L}{\mu_j})^{\frac{1}{2}}(x_j)^{\frac{1}{2}}$. Under spatial equilibrium, individual households' utilities are equalized across regions such that $u(c_j, n_j) = \overline{u}$. Given a realization of exogenous government purchase across regions $\{G_l\}_{l\in\mathcal{J}}$, the spatial equilibrium condition pins down the equilibrium distribution of population across regions, denoted by $\mu_j = \mu(G_j, \{G_l\}_{l\in\mathcal{J}})$. Thus, regional construction spending depends on the size of the regional population, denoted by $x_j = x(\mu_j) = \mu_j/L$, where $x_\mu > 0$ is the derivative of construction spending with respect to the size of regional population. The equilibrium regional labor supply function is given by $n_j = n(g_j, x_j) = \frac{\bar{c}+g_j+x_j}{a+\lambda_n/a}$, where $n_g > 0$ and $n_x > 0$ are the first-order derivative of labor supply with respect to government purchases and construction spending. Under spatial equilibrium, the regions receiving relatively high government purchases $G_j > G$ attract labor immigration such that $\mu_j > \mu^\star$, and vice versa.

The following proposition characterizes the equilibrium regional and aggregate output:

Proposition 1. *Under spatial equilibrium, the exact solution for regional output per resident in response to government purchase shocks is given by*

$$y_j = y^\star + an_g \cdot (g_j - g^\star) + an_x \cdot (x_j - x^\star) \tag{2.1}$$

and the exact solution for aggregate output is

$$Y = \sum_j y_j \mu_j = Y^\star + \underbrace{an_g \cdot (\sum_j G_j - G^\star)}_{\textit{mean effect}} + \underbrace{an_x \cdot \sum_j (x_j - x^\star)(\mu_j - \mu^\star)}_{\textit{covariance effect}}, \tag{2.2}$$

where regional per resident government purchases are $g_j = G_j/\mu_j$, construction spending is $x_j = \mu_j/L$, and equilibrium regional population is $\mu_j = \mu(G_j, \{G_l\}_{l\in\mathcal{J}})$.

As Proposition 1 states, equation (2.1) indicates how regional government-purchase shocks affect regional output per resident. First, regional government purchases have a direct stimulating effect on regional output, represented by $an_g(g_j - g^\star)$. Second, as labor immigration increases regional housing demand, regional construction spending booms and pushes regional output up additionally, reflected by $an_x(x_j - x^\star)$. This amplification mechanism triggered by asymmetric regional government-purchase shocks is different from the existing literature based on monetary union models with fixed populations across regions.

Aggregating equation (2.1) weighted by populations, the aggregate output can be represented by $Y = Y^\star + an_g(G - G^\star) + an_x(X - X^\star)$. Using a covariance decomposition, this yields $X - X^\star = \sum_j x^\star(\mu_j - \mu^\star) + \sum_j (x_j - x^\star)\mu^\star + \sum_j (x_j - x^\star)(\mu_j - \mu^\star)$. Due to the equilibrium construction policy $x_j = \mu_j/L$ and adding-up condition of population, the first two terms wash out. Therefore, equation (2.2) shows how aggregate output changes in response to asymmetric government purchase shocks across regions.

The "mean effect" term $an_g(\sum_j G_j - G^\star)$ represents the standard expansionary effect of government spending; it shows the effect of aggregate government spending $\sum_j G_j$ independent of its distribution across regions. This result is consistent with the existing literature. In the case without regional reallocation, the solution for aggregate output is given by $Y^o = Y^\star + an_g(\sum_j G_j - G^\star)$. It implies that the national effect of government spending only depends on the aggregate amount, because the effects of regional output in response to asymmetric regional government spending net out.

The novel mechanism is the "covariance effect" term $an_x \sum_j (x_j - x^\star)(\mu_j - \mu^\star)$, which represents the second-order effect arising from the comovement of regional construction and population. High-spending regions ($G_j > G^\star$) attract labor immigration ($\mu_j > \mu^\star$) and also have more construction demand ($x_j > x^\star$). Plugging in equilibrium construction spending, the covariance effect term becomes $an_x/L \cdot \sum_j (\mu_j - \mu^\star)^2 > 0$. Therefore, regional reallocation in response to asymmetric government spending shocks leads to a positive comovement between regional construction and population. This covariance effect amplifies aggregate output relative to a model without labor reallocation.

2.2.3 General Solution

Proposition 1 shows the regional reallocation mechanisms analytically for specific functional forms. In general, the regional reallocation mechanisms in equation (2.2) also appear in cases where the above functional form assumptions are relaxed. In this subsection, I show that although the macro impacts of regional reallocation wash out in the first-order approximation, they can be important in a second-order approximation.

Proposition 2. *Under spatial equilibrium, the first-order approximation of the response of regional output to government purchase shocks is given by*

$$y_j = y^\star + an_g \cdot (g_j - g^\star) + an_x(x_j - x^\star) + h.o.t \tag{2.3}$$

and the second-order approximation to aggregate output is given by

$$Y = Y^\star + an_g(\sum_j G_j - G^\star) + \underbrace{an_x \sum_j \tilde{x}_\mu(\mu_j - \mu^\star)^2}_{covariance\ effect} + \frac{1}{2}an_{gg}(\sum_j G_j - G^\star)^2 + h.o.t., \tag{2.4}$$

where $\tilde{x}_\mu \equiv (x_\mu + \frac{1}{2}x_{\mu\mu}\mu^\star) > 0$, x_μ and $x_{\mu\mu}$ are first- and second-order derivatives of construction spending with respect to regional population.

As Proposition 2 illustrates, the regional reallocation effects on regional output show up in the first-order approximation while the regional reallocation effects on aggregate output appear in the second-order approximation.

Consistent with equation (2.1), the construction effect $an_x(x_j - x^\star)$ in equation (2.3) amplifies regional output relative to the one with fixed populations across regions. This effect is first-order to regional output. As more labor immigrates to a region due to favorable regional shocks, regional construction spending booms in response and increases regional output.

The aggregate impacts of regional reallocation are shown by the covariance effect term $\sum_j \tilde{x}_\mu(\mu_j - \mu^\star)^2$ in equation (2.4). Net directed migration shifts population to booming regions which are where populations and regional output per resident go up jointly. This comovement force generates a nonlinear aggregate effect that amplifies the impacts of government spending relative to the situation without labor migration.

Such nonlinear aggregate effects are similar to the generalized Hulten Theorem in

Baqaee and Farhi (2018) in the following sense. Given a fixed amount of aggregate government spending, the regional output effects of asymmetric government-spending shocks net out in the aggregate to the first-order approximation. However, the second-order aggregate effect of a regional shock is associated with the change in the population share of that region. In response to a positive regional government-spending shock, not only the regional output per resident rises but also the population share in that region. Regional reallocation amplifies the aggregate impact of positive regional shocks and mitigates that of negative regional shocks, leading to an amplification of aggregate output impact to a second-order approximation.

Under a first-order approximation, the aggregate impacts of regional reallocation wash out. The standard log-linearization approach cannot deliver the macro implications of regional reallocation in response to asymmetric shocks across regions. Therefore, in Section 2.4, I propose a flexible perturbation method to characterize the macro impacts of regional reallocation in a tractable manner.

2.3 A Quantitative Multi-Region New Keynesian Model

In this section, I develop a multi-region New Keynesian model with imperfect labor mobility and housing construction to quantify the role of regional reallocation in affecting regional and aggregate fiscal multipliers. This model is built on a large literature on monetary union models.[7] The economy consists of multiple regions indexed by $j \in \mathcal{J} = \{1, ..., J\}$, where capital and intermediate goods are freely traded. The economy has a federal government that conducts fiscal and monetary policy.[8] However, I relax the assumption that all regions are small open economies

[7]For example, Benigno and Benigno (2003), Gali and Monacelli (2008), NS (2014), and Beraja, Hurst and Ospina (2018).

[8]The regional reallocation mechanism does not depend on nominal rigidity, as shown by the simple model in Section 2.2. The purpose of introducing the New Keynesian framework is to incorporate

with fixed populations and allow individuals to relocate from one region to another subject to migration frictions. Individuals decide optimal migration policies subject to a constant Calvo probability of being able to relocate and idiosyncratic taste shocks for various regions. In addition, variable construction completion lags help to generate hump-shaped responses of construction as in the data. In this model, subscripts j and t denote region and period and lower-case letters for regional per resident terms and upper-case letters for total quantities or prices.

2.3.1 Firms and Production

2.3.1.1 Local Production

Regional Final Goods. Each region has a continuum of competitive firms that purchase traded intermediate goods and regional labor to produce regional final goods. Each firm operates a Cobb-Douglas production function, $y_{jt} = Am_{jt}^{\alpha}n_{jt}^{1-\alpha}$, where y_{jt}, m_{jt}, and n_{jt} denote the amount of regional output, intermediate input, and hours worked in per resident terms, and A the productivity level.[9] A regional final goods firm chooses an input bundle to maximize its profit,

$$\max_{\{m_{jt},n_{jt}\}} : P_{jt}y_{jt} - W_{jt}n_{jt} - Q_t m_{jt},$$

where P_{jt} and W_{jt} are the price of regional final goods and the nominal wage in region j at time t, and Q_t is the price of traded intermediate inputs. The optimal input decisions for regional final goods firms satisfy $W_{jt} = (1-\alpha)P_{jt}Am_{jt}^{\alpha}n_{jt}^{-\alpha}$ and $Q_t = \alpha P_{jt}Am_{jt}^{\alpha-1}n_{jt}^{1-\alpha}$.

monetary policy reaction to generate realistic aggregate fiscal multipliers.

[9] The size is normalized as being the same as regional population. The regional total output of final goods is denoted by $Y_{jt} = y_{jt}\mu_{jt}$

Regional Housing Construction. In each period a fraction δ_h of housing stock depreciates such that the regional housing stock evolves as $H_{jt+1} = (1-\delta_h)H_{jt} + Y^h_{jt}$, where Y^h_{jt} denotes the number of new houses built at time t.[10] Assume that there is a continuum of competitive builders whose size is normalized by regional population. Each builder operates a decreasing return-to-scale housing production technology, $y^h_{jt} = bx^{\zeta}_{jt}$, where $\zeta \in (0,1)$ and x_{jt} and b are the amount of regional goods used as construction inputs and the productivity of the housing sector.[11] The total amount of newly-built houses is $Y^h_{jt} = y^h_{jt}\mu_{jt}$.

In order to generate hump-shaped responses of regional construction, the model needs to introduce some frictions on regional housing supply. Permit restrictions are a natural friction. In practice, the government issues a limited number of permits to restrict the regional supply of housing due to political, environmental, and other reasons. In this model, I assume that the government issues one unit of housing permit for each new immigrant.[12] I also assume that a fraction ϕ_h of the authorized permit stock S_{jt} is completed in each period, where $\phi_h \in (0,1)$ denotes the completion rate. Thus, the number of new houses built at time t is not greater than the number of completed permits, $Y^h_{jt} \leq \phi_h S_{jt}$. The stock of authorized permits S_{jt} is the sum of the unfinished permits for the last period $(1-\phi_h)S_{jt-1}$ and new issued permits S^n_{jt}, given by $S_{jt} = (1-\phi_h)S_{jt-1} + S^n_{jt}$.[13] Under the specific completion rate, the average

[10] In this quantitative model, I allow for varying housing consumption, instead of inelastic demand. More details about utility specification and regional housing market structure are in subsection 2.3.2.

[11] I also study the case of a competitive regional builder who operates a housing production function of $Y^h_{jt} = bX^{\zeta}_{jt}$. Most of the results are preserved. But the regional steady state interacts with the size of regional population, and it makes aggregation relatively more difficult than the above specification of the per resident production function.

[12] Howard and Liebersohn (2018) establish an empirical fact that the change of population is nearly one-for-one with the change of housing units, implying that labor reallocation is the main force for cross-sectional housing demand. The "one permit per person" assumption is consistent with this empirical fact.

[13] Although regional population is a constant in steady state, there is still an influx of new immigrants, which exactly offsets the amount of labor out-migration, due to idiosyncratic location-taste shocks. Therefore, at steady state, new permits are issued to cover the depreciated amount of

duration of a construction project is $\frac{1-\phi_h}{\phi_h}$ periods. This parameter can be disciplined by the average construction duration from the data.[14]

Therefore, builders optimally choose the number of houses to build subject to permit restrictions. The equilibrium rental rates and housing prices are pinned down by the individual household's marginal rate of substitution between consumption and housing and the user cost equation.[15]

2.3.1.2 Production of Tradeable Factors

Monopolistic Intermediate Firms. For each variety i (denoted by parentheses), a monopolistic firm uses capital to produce intermediate goods by $M_t(i) = K_t(i)$.[16] Since the intermediate goods and capital goods are freely traded, the location of intermediate goods production does not matter. Each monopolistic firm is subject to Calvo-type nominal rigidity such that it resets its price each period with a constant probability $1-\theta$. Therefore, the profit-maximization problem is $\max_{\{Q_t^\star\}} : \mathbb{E}_t \left\{ \sum_{s=0}^{\infty} \theta^s \Lambda_{t,t+s}^n [Q_t^\star M_{t+s|t}(i) - \Phi_{t+s}^n M_{t+s|t}(i)] \right\}$, subject to its demand, $M_{t+s|t}(i) = [Q_t^\star / Q_{t+s}]^{-\varepsilon} M_{t+s}$ for all $s \geq 0$, where $Q_t^\star$ is the optimal reset price for the firm that is able to reset its price at time t, $\Lambda_{t,t+s}^n$ and Φ_t^n denote nominal stochastic discount factor and nominal marginal cost at time t. Thus, the optimal reset price satisfies the optimal pricing condition, $\mathbb{E}_t \left\{ \sum_{s=0}^{\infty} \theta^s \Lambda_{t,t+s}^n M_{t+s|t}(i) \left(Q_t^\star - \frac{\varepsilon}{\varepsilon-1} \Phi_{t+s}^n \right) \right\} = 0$, implying that a firm resets its price equal to a constant markup over a weighted average of current and expected future nominal marginal costs. The nominal marginal

housing stock.

[14] If ϕ_h goes to one, all housing permits are completed immediately and the model degenerates to a standard housing model without completion duration.

[15] The builder's optimality conditions as well as equilibrium conditions for rental rate and housing prices are shown in the Appendix.

[16] For simplification, I specify a constant return-to-scale production function of intermediate goods. In fact, the production function can be also decreasing return-to-scale, e.g., $M_t(i) = [K_t(i)]^\kappa$, where $\kappa \in (0,1)$.

cost of intermediate producers equals to the rental rate of capital, $\Phi_t^n = R_t^k$. Given the Calvo-probability specification, the price of traded intermediate goods follows $Q_t = \left[\theta Q_{t-1}^{1-\varepsilon} + (1-\theta) Q_t^{\star 1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}$.

Traded intermediate goods M_t are produced by a constant elasticity of substitution (CES) production function over a continuum of varieties indexed by $i \in [0,1]$, $M_t = [\int_0^1 M_t(i)^{\frac{\varepsilon-1}{\varepsilon}} di]^{\frac{\varepsilon}{\varepsilon-1}}$, where ε governs the substitutability of varieties with $\varepsilon \geq 1$. Hence, the price of traded intermediate goods is and the demand for intermediate goods of variety i is $M_t(i) = [Q_t(i)/Q_t]^{-\varepsilon} M_t$.

Investment Goods Producer. A competitive investment goods producer combines regional final goods to produce investment goods using a CES production function, $I_t = (\sum_j i_{jt}^{\frac{\eta-1}{\eta}} \mu_{jt})^{\frac{\eta}{\eta-1}}$, where η governs the substitutability of regional goods with $\eta \geq 0$ and μ_{jt} denotes the population in region j at time t. Hence, the demand for local goods in region j to produce investment goods is $i_{jt} = (P_{jt}/P_t)^{-\eta} I_t$ and market price of investment goods (and capital) is $P_t = (\sum_j P_{jt}^{1-\eta} \mu_{jt})^{\frac{1}{1-\eta}}$. Therefore, the aggregate capital stock evolves according to $K_{t+1} = (1-\delta_k)K_t + I_t$, where δ_k is the capital depreciation rate.

2.3.2 Households

There is a continuum of individual households in the economy. The size of a population is normalized to one. Each household lives infinitely, and time is discrete $t = 1, 2, ...$ Households are allowed to migrate across regions but are subject to frictions. Within each period individuals have to live in the same region where they work. Individuals hold equal shares of capital stock and receive dividend payoffs. They decide consumption c_{jt}, hours worked n_{jt}, housing service h_{jt}, and migration choices to maximize their expected lifetime utility given by $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_{jt}, n_{jt}, h_{jt})$, where β is a time discount factor. I also assume a big family who decides the aggregate stock of capital. This assumption enables me to avoid keeping track of capital distribution across individuals as a state variable.

2.3.2.1 Imperfect Labor Mobility across Regions

I now introduce two features to generate imperfect labor mobility across regions—Calvo migration probability and Gumbel location-preference shocks. At the end of each period, each individual agent can relocate with a constant probability $\psi \in (0, 1)$, which is referred to as the Calvo migration probability. In addition, following the standard specification in the spatial economics literature (Glaeser and Gottlieb, 2009; Moretti, 2011), individuals who relocate face idiosyncratic location-specific preference shocks, ϵ_{jt}, which follow an i.i.d. Gumbel distribution across individuals and over time.[17] The evolution of population distribution across regions is given by

$$\mu_{jt+1} = (1 - \psi)\mu_{jt} + \psi\varphi_{jt}, \tag{2.5}$$

where μ_{jt} denotes the population in region j at time t and φ_{jt} stands for the optimal migration policy to region j in the next period for those who are allowed to relocate.[18]

2.3.2.2 The Individual Household's Problem

Each individual household holds equal shares of capital and therefore receives dividends from a parent. In addition, each individual pays housing rent to the parent to obtain housing service. Given that households receive local real wage income, profit

[17] The Gumbel distribution assumption is standard in dynamic discrete choice models and allows for simple aggregation of idiosyncratic decisions made by individual households. (Artuc, Chaudhuri, and McLaren, 2010; Caliendo, Dvorkin, and Parro, 2018)

[18] The newly issued permits at time t is $S^n_{jt} = \psi\varphi_{jt}$, under the one permit per immigrant assumption.

from monopolistic firms R_t^m, and dividend D_t, they decide consumption c_{jt}, hours worked n_{jt}, housing service h_{jt}, and migration choices, to maximize expected lifetime utility

$$V_j(\boldsymbol{s}) = \max_{\{c_{jt}, n_{jt}, h_{jt}\}} \left\{ u(c_{jt}, n_{jt}, h_{jt}) + \kappa_j + \beta[(1-\psi)\mathbb{E}_t V_j(\boldsymbol{s}') + \psi(\max_{l \in \mathcal{J}} : \mathbb{E}_t V_l(\boldsymbol{s}') + \nu\epsilon_{lt})] \right\},$$

subject to budget constraint, $c_{jt} + \varrho_{jt}^h h_{jt} = (1-\tau_t)(W_{jt}/P_{jt})n_{jt} + D_t + R_t^m - T_t$, where $\boldsymbol{s}$ denotes a set of state variables at time t, κ_j denotes local amenity in region j,[19] ψ is the Calvo migration probability, ϱ_{jt}^h is the market rental rate of housing, τ_t and T_t are labor income tax rates and real lump-sum taxes levied by a federal government, and ϵ_{jt} denotes idiosyncratic location-preference shocks, and ν is the standard deviation of location-preference shocks.

Under the Gumbel distribution assumption, the optimal relocation policy is shown by[20]

$$\varphi_{jt} = \frac{\exp[\mathbb{E}_t V_j(\mathbf{s}_{t+1})]^{\frac{1}{\nu}}}{\sum_l \exp[\mathbb{E}_t V_l(\mathbf{s}_{t+1})]^{\frac{1}{\nu}}}. \tag{2.6}$$

This optimal relocation policy is consistent with the existing literature of labor reallocation (e.g., Artuc, Chaudhuri, and McLaren, 2010; Caliendo, Dvorkin, and Parro, 2018). First, the optimal migration policy is a probabilistic rule due to idiosyncratic location-preference shocks. Second, the regions with higher virtual expected value $\mathbb{E}_t V_j(\mathbf{s}_{t+1})$ attract a larger fraction of labor migration. Third, a higher standard deviation of location-preference shocks ν implies a flatter distribution of optimal relocation φ_t. Intuitively, the optimal migration policy is determined by two forces—the gain of expected value from relocation and the cost from location-preference shocks.

[19] The region-specific time-invariant local amenity is commonly assumed in the literature on spatial economics (see Redding and Rossi-Hansberg (2017) for a survey. The local amenity term represents climate, landscape, culture, language, etc.)

[20] The proof is in the Appendix.

A higher variance of location preference shocks causes higher expected loss to a particular location.

2.3.2.3 The Parent's Problem and Capital Accumulation

A representative "parent" decides aggregate capital stock K_{t+1}, nominal bond holding B_t, and dividend D_t to maximize the average of all individual households' utilities, $\max_{\{K_{t+1},B_{t+1}\}} : \mathbb{E}_0 \sum_t \beta^t \left(\sum_j u(c_{jt}, n_{jt}, h_{jt})\mu_{jt}\right)$ subject to the budget constraint

$$P_t K_{t+1} + B_t = R_{t-1}B_{t-1} + (1-\delta_k)P_t K_t + R_t^k K_t - \sum_j P_{jt} D_t \mu_{jt}$$
$$+ \sum_j P_{jt} \varrho_{jt}^h h_{jt}\mu_{jt} - \sum_j P_{jt} x_{jt}\mu_{jt},$$

where R_t is nominal interest rate, R_t^k is nominal return rate of capital, and $\sum_j P_{jt}\chi_{jt}\mu_{jt}$ is the total spending on housing construction. Bonds are in zero net supply in equilibrium such that $B_t = 0$ for all t. The parent receives all profit from housing builders, owns all the houses, and rents them to individual households. This representative parent assumption enables me to avoid keeping track of the distribution of capital holding across individuals as a state variable.[21] The intertemporal Euler equation to the parent's optimization problem is given by $MU_t = \beta \mathbb{E}_t \left[R_t \frac{P_t}{P_{t+1}} MU_{t+1}\right]$, where $MU_t = \sum_j u_c(jt)\mu_{jt}$ represents economy-wide marginal utilities and the nominal stochastic discount factor is $\Lambda_{t,t+s}^n = \frac{P_t MU_{t+s}}{P_{t+s} MU_t}$.

[21] The big family specification is not the same as complete market assumption. Although households have perfect risk sharing within a region, their consumption could be different across regions in the model.

2.3.3 Government

2.3.3.1 Region-based Government Purchases

The economy has a federal government that conducts fiscal and monetary policy. The per-resident federal government spending in each region $\{g_{jt}\}$ has the common component ξ_t and a regional idiosyncratic component z_{jt} over regional final goods such that $g_{jt} = \exp(\xi_t + z_{jt})$. The common component is assumed to be a constant $\xi_t = \overline{\xi}$ and idiosyncratic components follow exogenous AR(1) processes given by $z_{jt} = \rho_z z_{jt-1} + \epsilon^z_{jt}$, where ρ_z denotes the persistence of idiosyncratic government spending and ϵ^z_{jt} is an i.i.d. white noise with zero mean. The government spending at steady state is $\overline{g} = \exp(\overline{\xi})$.

2.3.3.2 National Tax and Monetary Policy

The government levies both labor income and lump-sum taxes to finance its purchases of goods. Therefore, the government budget constraint satisfies

$$\sum_j P_{jt} g_{jt} \mu_{jt} = (1 - \tau_t) \sum_j W_{jt} n_{jt} \mu_{jt} + \sum_j P_{jt} T_t \mu_{jt}.$$

The federal government also operates a common monetary policy for the whole economy. In particular, I consider a Taylor rule for nominal interest rates,

$$\hat{R}_t = \rho_r \hat{R}_{t-1} + (1 - \rho_r)(\phi_\pi \hat{\pi}_t + \phi_y \hat{Y}_t^{gap}),$$

where $\hat{R}_t$, $\hat{\pi}_t$, and $\hat{Y}_t^{gap}$ denote the log-deviations of the nominal interest rate, inflation rate, and aggregate output gap, respectively. ρ_r, ρ_π, and ρ_y stand for the policy responsiveness to interest rate in the last period, inflation rate, and aggregate output gap. Aggregate output Y_t is defined by summing up the value of all final goods across regions, $Y_t \equiv \sum_j P_{jt} y_{jt} \mu_{jt} / P_t$.

2.3.4 Equilibrium

Definition. *A dynamic spatial equilibrium is a collection of prices and quantities,* $\{P_{jt}, W_{jt}, Q^h_{jt}, \varrho^h_{jt}, Q_t, P_t, Q^\star_t, R_t, R^k_t, Q_t(i)\}$ *and* $\{c_{jt}, m_{jt}, n_{jt}, y_{jt}, \mu_{jt}, g_{jt}, x_{jt}, h_{jt}, i_{jt}, \varphi_{jt}, K_t, D_t, I_t, T_t, M_t, M_t(i), K_t(i)\}$ *for each region and period, and exogenous shocks of government spending across regions* $\{z_{jt}\}$*, such that (i) regional final goods firms, monopolistic intermediate firms, investment goods producers, and housing builders maximize profits; (ii) individual households maximize lifetime expected utility; (iii) the parent maximizes weighted average utility; (iv) regional final goods, labor, and housing markets and national intermediate good and capital markets clear such that* $y_{jt} = c_{jt} + g_{jt} + i_{jt} + x_{jt}$ *for all regions,* $M_t = \sum_j m_{jt}\mu_{jt}$ *and* $K_t = \int_0^1 K_t(i)di$ *for all time.*

2.4 Aggregation

In this section, I develop a new method to characterize and solve the regional and aggregate dynamics in the model. I show that the equilibrium behavior of regional and aggregate variables in a locally-perturbed economy can be represented by a regional dynamic system and an aggregate dynamic system, respectively. Applying this useful aggregation result, I provide a procedure to solve the model.

2.4.1 Aggregation

The model is log-linearized around the steady state. A regional variable's log-deviation from steady state is denoted by a lowercase letter with hats. Uppercase letters with hats in calligraphic form represent aggregate variables in log-deviation from steady state. For example, $\hat{c}_{jt} \equiv \log(c_{jt}/c^\star)$ and $\hat{\mathcal{C}}_t \equiv \log(C_t/c^\star)$ stand for regional per resident and aggregate consumption deviations from the steady state, where $c^\star$ is the steady state for per-resident consumption. In addition, the difference of log-deviation of a regional variable from another regional counterpart is denoted by a lowercase letter with tildes, e.g., $\tilde{c}_{jt} \equiv \hat{c}_{jt} - \hat{c}_{j't}$.

As shown in Section 2.2, the aggregate effects of regional reallocation totally wash out when using a first-order approximation. To capture the covariance term arising from regional reallocation, I apply a flexible perturbation method. For any aggregate variable that is defined by the sum of per-resident regional counterparts weighted by regional populations, e.g., aggregate consumption $C_t = \sum_j c_{jt}\mu_{jt}$, and due to the fact that population shares across regions add up to one $\sum_j \mu_{jt} = 1$, we have the following identity, $\frac{C_t - c^\star}{c^\star} = \frac{\sum_j c_{jt}\mu_{jt} - \sum_j c^\star \mu_{jt}}{c^\star} = \Sigma_j(\frac{c_{jt} - c^\star}{c^\star})\mu_{jt}$. Therefore, $\hat{\mathcal{C}}_t = \sum_j \hat{c}_{jt}\mu_{jt}$, implying that the cross-sectional inner-products between local variable exposures and population shares are equivalent to the aggregate counterparts in log-deviation from steady state. In addition, expanding μ_{jt} around steady state, i.e., $\mu_{jt} = (1 + \hat{\mu}_{jt})\mu_j^\star$, the log-deviation of aggregate consumption can be shown by $\hat{\mathcal{C}}_t = \sum_j \hat{c}_{jt}\mu_j^\star + \sum_j \hat{c}_{jt}\hat{\mu}_{jt}\mu_j^\star$. This equation indicates that the second-order term $\sum_j \hat{c}_{jt}\hat{\mu}_{jt}\mu_j^\star$ captures the covariance effect arising from regional reallocation. The major advantage of such a flexible perturbation method is that it enables us to characterize the nonlinear covariance effect in a linear manner, which significantly reduces computational intensity but preserves the covariance effect.

Applying this flexible perturbation method, I aggregate up all log-linearized equilibrium conditions by population shares. Under two key assumptions, the model can aggregate up to an augmented representative economy that is able to capture the impacts of cross-sectional distribution on aggregate dynamics. Therefore, the aggregate dynamics in the model—the equilibrium behavior of aggregate variables in log-deviation from steady state—can be characterized by such a log-linearized augmented representative economy.

The first key assumption for aggregation is that the underlying production, utility, and market structure are symmetric across regions. For instance, individual households have the same utility form over consumption and leisure and are linear to the exogenous local amenity. The second assumption is the Gumbel specification of location-preference shocks. This assumption implies an optimal migration policy that only depends on the differentials of residential value across regions.

Under these assumptions, in steady state each region is identical in per resident terms. Given the Gumbel specification, heterogeneous local amenities only affect the size of population shares across regions at steady state. Therefore, the log-linearized aggregate variables are equivalently represented by the inner-product of log-linearized regional variables and varying regional population shares, e.g., $\hat{\mathcal{C}}_t = \sum_j \hat{c}_{jt}\mu_{jt}$ for aggregate consumption. Similarly, I define the inner-product of a log-linearized regional variable and the difference between optimal migration policy and population distribution $\tilde{\varphi}_{jt} \equiv \varphi_{jt} - \mu_{jt}$ as a "reallocation summary statistic," denoted by an uppercase letter with tildes, e.g., $\tilde{\mathcal{G}}_t \equiv \sum_j \hat{g}_{jt}\tilde{\varphi}_{jt}$ for localized government spending. The reallocation summary statistic characterizes how regional reallocation affects aggregate dynamics in response to regional economic conditions. Thus, Proposition 3 shows that the aggregate dynamics in the model can be characterized by a log-linearized augmented representative aggregate economy.

Proposition 3 (*Aggregate Dynamics*) *The behavior of aggregate variables in the locally-perturbed economy is identical to that of an augmented representative economy in the form of a linear dynamic system* $\Gamma_0 \mathbb{E}_t \hat{\mathcal{Y}}_{t+1} = \Gamma_1 \hat{\mathcal{Y}}_t + \Gamma_2 \tilde{\mathcal{X}}_t$*, where* Γ_0*,* Γ_1*, and*

Γ_2 *are matrices of coefficients,* $\hat{\mathcal{Y}}_t = \sum_j \hat{\mathrm{y}}_{jt}\mu_{jt}$ *stands for aggregate variables and* $\tilde{\mathcal{X}}_t = \sum_j \hat{\mathrm{x}}_{jt}\tilde{\varphi}_{jt}$ *stands for reallocation summary statistics.*

Proof. See Appendix A.

Proposition 3 clearly shows the aggregate influence from regional reallocation. Under the Gumbel specification, the optimal migration policy $\{\varphi_{jt}\}$ depends on the distribution of regional economic conditions $\{\hat{\mathrm{x}}_{jt}\}$. Therefore, the reallocation summary statistics $\tilde{\mathcal{X}}_t = \sum_j \hat{\mathrm{x}}_{jt}\tilde{\varphi}_{jt}$ are able to characterize the aggregate influence of asymmetric regional shocks through regional reallocation. For instance, denote a regional demand by $\hat{\mathrm{x}}_{jt}$. Holding other regions fixed, a positive unit of local demand shock $\hat{\mathrm{x}}_{jt} > 0$ attracts more labor immigration $\tilde{\varphi}_{jt} > 0$. Therefore, the augmented aggregate demand rises, $\tilde{\mathcal{X}}_t > 0$, due to the covariance effect. In contrast, without net directed labor reallocation, $\tilde{\varphi}_{jt}$ term vanishes and therefore aggregate dynamics has nothing to do with regional reallocation and is irrelevant to the cross-sectional distribution of regional shocks.

Subtracting the aggregate system from the original log-linearized system, this yields a regional dynamic system that contains log-deviation of regional variables from the aggregate counterparts. Proposition 4 shows that the regional dynamics in the model can be characterized by a regional dynamic system.

Proposition 4 (*Regional Dynamics*) *The behavior of regional variables in the log-linearized economy is identical to that of a regional dynamic economy in the form of* $\Phi_0\mathbb{E}_t\tilde{\mathbf{y}}_{t+1} = \Phi_1\tilde{\mathbf{y}}_t + \Phi_2\tilde{\mathbf{z}}_t$*, where* Φ_0*,* Φ_1*, and* Φ_2 *are matrices of coefficients,* $\tilde{\mathbf{y}}_t$ *stands for differential regional variables,* $\tilde{\mathbf{z}}_t$ *represents regional idiosyncratic shocks.*

Proof. See Appendix A.

Proposition 4 provides two useful results. First, the solution for regional dynamics is directly comparable with the estimated regional impulse responses by a local projection method for panel data with time fixed effects. The time-fixed effects in regression soak up aggregate variation such as national monetary shocks or aggregate productivity. The differential regional variables in log-deviation from steady state also net out all aggregate variation. Therefore, relevant parameters can be calibrated by matching the regional impulse responses in the model with the estimated regional impulse responses from the data. Second, given the realization of localized government spending shocks across regions, the optimal migration policy can be solved, as well as the evolution of population distribution across regions. Under the Gumbel specification of location-preference shocks, the optimal migration policy only depends on the differentials of regional economic conditions. All common changes of aggregate conditions play no role in affecting labor reallocation across regions.

2.4.2 Solution Method

Numerically solving the equilibrium of a spatial DSGE model is challenging. Migration decisions depend on regional economic conditions (e.g., regional population density and regional fiscal shocks) not only in an individual's residential region but also in all other regions (i.e., all 51 regions). Thus the entire cross-sectional distribution of regional economic conditions evolves as a state variable, causing a computational complexity.

Applying the aggregation results, I solve the aggregate dynamics using locally accurate perturbation methods, thus avoiding computational complexity of high dimensionality. I first solve the regional dynamics and obtain optimal labor migration policies and the evolution of labor distribution across regions. Then, given the distribution of regional demand shocks, I employ the solution for labor migration and solve for the evolution of reallocation summary statistics. Finally, I plug their evolution processes into the aggregate dynamic system and solve for aggregate impulse responses.

This approach is similar to the method of Beraja, Hurst, and Ospina (2018) in that it enables us to separately solve regional and aggregate dynamics by using perturbation methods. Unlike their method, my approach allows for labor migration across regions, and aggregate variables can depend on the endogenously varying population shares. Therefore, the evolution of aggregate shocks depends not only on the exogenous shock process but also on the cross-sectional labor distribution in response to the shocks.

2.5 Model Parameterization

The model is calibrated in two steps. First, a set of parameters is fixed to be consistent with the standard macroeconomic literature. Given those parameters, the remaining parameters are calibrated to match the estimated local impulse responses to fiscal stimuli from Chapter 1 and related moments from the data.

Fixed Parameters. Table (2.1) lists the parameters to be fixed exogenously. Household's utility is separable over consumption, hours worked, and housing service, $u(c,n,h) = \log(c) - \lambda_n \frac{n^{1+\phi_n^{-1}}}{1+\phi_n^{-1}} + \lambda_h \log(h)$, where ϕ_n denotes the Frisch elasticity of labor supply.[22] As in NS (2014), ϕ_n is set to 1, which is relatively standard in macroeconomics, although ϕ_n is somewhat higher than values estimated in microeconomic studies (Hall, 2009; Chetty *et al.*, 2012). The calibrated value of λ_n is pinned down by setting the average hours worked $N^\star = 0.2$ in steady state.[23] I calibrate $\lambda_h = 0.22$ to

[22] The major results are robust to a CRRA utility function, as shown in the Appendix.

[23] The optimal consumption-labor condition at steady state, $\frac{W^\star}{P^\star c^\star} = \lambda_n (N^\star)^{\phi_n}$, can pin down the value of $\lambda_n = 1.035$ by setting average hours worked ratio $N^\star = 0.2$.

match the average ratio of housing consumption over total consumption, 18 percent. (Piazzesi and Schneider, 2016)

A model period is one year. The time discount factor is set to $\beta = 0.97$, implying that the annual real interest rate at steady state is 3 percent. The annual capital depreciation rate is set to $\delta_k = 0.08$, which is roughly in line with the average in the postwar data, and annual housing depreciation rate to $\delta_h = 0.03$ consistent with Piazzesi and Schneider (2016). According to NS's estimates of the persistence of government spending, the AR(1) coefficient is calibrated to be $\rho_z = 0.75$ at an annual frequency. The firm's annualized probability of resetting prices equals to $1 - \theta = 0.7$, which is in line with the fact that firms adjust their prices with probability 0.25 in each quarter on average (Bils and Klenow, 2004).

The regional final good production is of Cobb-Douglas form with a labor share $1 - \alpha = \frac{2}{3}$. The elasticity of substitution across intermediate varieties is specified to $\varepsilon = 7$ as in NS. The elasticity of substitution across regions is set to be $\eta = 1$. This parameter is associated with the elasticity of substitution between home goods and foreign goods in the open economy macroeconomics literature, which ranges between 1 and 2 (Chari, Kehoe, and McGrattan, 2002). A larger η leads to more reduction in local demand from investment-goods production in response to local fiscal shocks.

In the benchmark case, the monetary policy follows a simple Taylor rule such that $\hat{r}_t = \phi_\pi \hat{\pi}_t$ with $\phi_\pi = 1.5$. This specification implies that the central bank aggressively raises the real interest rate to curtail the inflationary effects of a government spending shock. The benchmark case is the one in which government spending is completely financed by lump-sum taxes T_t. The ratio of federal government purchases over output is set to 10 percent, i.e., $\overline{g} = g^\star / y^\star = 0.1$, in steady state.

Parameter	**Description**	**Value**	**Source**
β	Annual discount factor	0.97	Standard
ϕ_n	Frisch elasticity	1	NS (2014)
λ_h	Housing utility coefficient	0.22	PS (2016)
$1-\alpha$	Labor share	0.67	NS (2014)
δ_k	Capital depreciation	0.08	Standard
δ_h	Housing depreciation	0.03	PS (2016)
θ	Annualized reset price probability	0.7	BK (2004)
ε	Elasticity of substitution across varieties	7	NS (2014)
η	Elasticity of substitution across regions	1	CKM (2002)
ϕ_π	Monetary policy responsiveness	1.5	Standard
ρ_z	Persistence of military spending	0.75	NS (2014)
$\overline{g}$	Steady state government purchase ratio	0.1	Standard

Table 2.1: Fixed Parameter Values

Notes: Parameters exogenously fixed in the calibration.

Fitted Parameters. I choose the remaining parameters to match impulse responses. Setting equal weights over horizons, I calibrate the parameters by matching model-based impulse responses and the estimated regional impulse responses of population and construction output from Chapter 1. The fitted parameters are summarized in Table 2.2.

The Calvo migration probability ψ is associated with the gross migration rate. The model with larger ψ gives rise to a higher gross migration rate. Thus, ψ is calibrated to match an annualized gross migration rate across states in the postwar United States of 3 percent from Molloy, Smith, and Wozniak (2011). The Gumbel elasticity ν governs the magnitude of net directed migration responses. The estimated regional impulse response of population can be used to discipline the Gumbel elasticity. The estimated regional impulse response of construction output is able to calibrate the permit completion rate ϕ_h and the housing production elasticity ζ. Since housing productivity b determines the construction spending in steady state, the average weight of regional output in the construction sector over the sample period, 5 percent,

Parameter	Description	Value
ψ	Calvo migration probability	0.031
ν	Gumbel elasticity	1.65
ϕ_h	Permit completion rate	0.47
ζ	Elasticity of housing production	0.74
b	Housing productivity	0.66

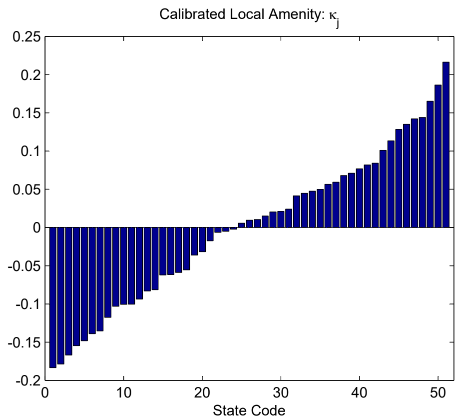
Table 2.2: Fitted Parameter Values

Notes: The targeted moments are, respectively, the annual gross migration rate across states for ψ, the estimated regional impulse responses to population for ν, the estimated regional impulse responses to construction output for ϕ_h and ζ, and the average weight of regional output in the construction sector for b.

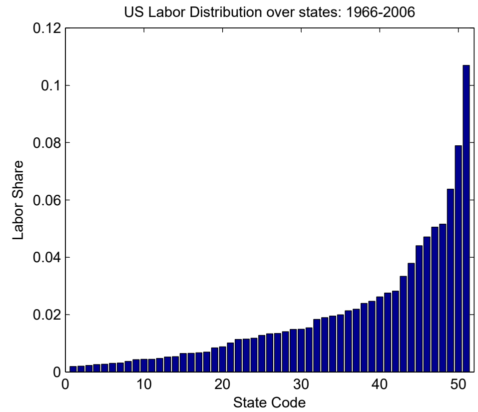
is a moment to pin down b. The model implies that the population distribution at steady state is determined by the local amenity parameters $\{\kappa_j\}$, and the average population shares by state over the sample period can discipline these parameters. The distributions of calibrated local amenity and population shares across states are reported by Figure 2.1.

The calibrated model fits the targeted moments well. Under these parameter values, the calibrated model generates a 3 percent simulated gross migration rate and 5 percent regional construction output ratio at steady state. The calibrated Gumbel elasticity $\nu = 1.65$ is broadly comparable to previous findings in the literature, e.g., Artuc, Cahaudhuri and McLaren's (2010) estimate $\nu = 1.88$ at an annual frequency. The calibrated housing production elasticity of non-land factors $\zeta = 0.74$ is close to the estimate in Epple, Gordon, and Sieg (2010), 0.85. The calibrated permit completion rate $\phi_h = 0.47$ implies the average duration of housing construction to be 1.1 years, which is consistent with the data provided by the Value of Construction Put in Place Survey from the Census.[24] Without labor reallocation across regions,

[24] According to the Value of Construction Put in Place Survey, a construction project, on average, takes one month for permit authorization, one month from authorization to start, and nine months from start to completion.



(a) Calibrated Amenity

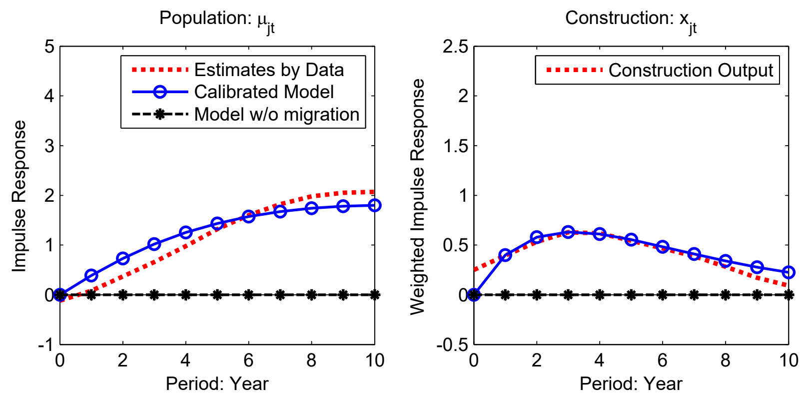


(b) Population Share

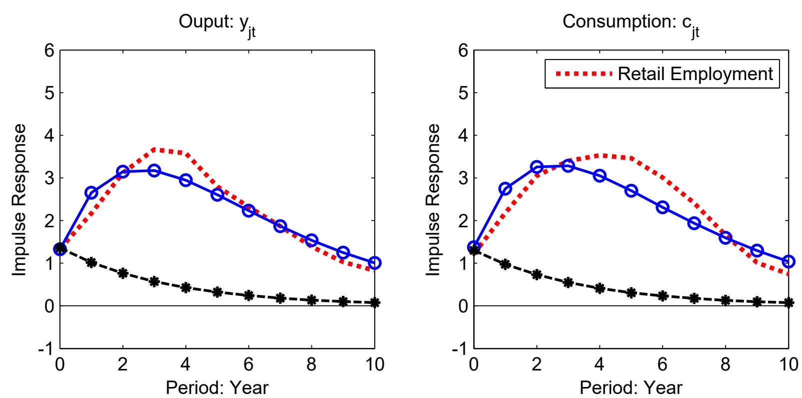
Figure 2.1: Distributions of Calibrated Amenity and Population across States
Notes: Panel (a) shows the calibrated value of local amenities across states, and Panel (b) displays the average population distribution across U.S. states from 1966 to 2006.

the model generates no responses of construction output and population, as shown in the Panel (a) in Figure 2.2.

Under the calibrated parameters, is the model able to replicate the identified impulse responses of regional economic activity to a military spending shock? Besides regional output per resident as a major measure of regional production, regional retail employment per capita can be used to proxy for regional consumption. As Guren *et al.* (2018) point out, retail employment has long been viewed by measurement agencies as one of the best available proxies for consumption expenditures (although not one-for-one). Panel (b) in Figure 2.2 shows that the calibrated model broadly matches the estimated regional impulse responses of a fiscal shock to output and consumption. In addition, without labor reallocation across regions, the model cannot generate the responses of output and consumption estimated from the data.



(a) Targeted Impulse Response: Population and Construction



(b) Untargeted Impulse Response: Output and Consumption

Figure 2.2: Matching Regional Impulse Responses in the Calibration

Notes: Red dotted lines are the estimated regional impulse responses from the data. Blue circles are the model-based regional impulse responses in the benchmark model, while black stars are the one in a model without labor migration. "Weighted impulse response" is the response of a variable weighted by its share in output. The construction share in output is 5 percent.

2.6 Quantitative Analysis of Regional Reallocation

Having shown that the calibrated model can generate regional dynamics that match the data, I now inspect the underlying mechanisms which drive these results. To this end, I study the quantitative impacts of a localized government spending shock by comparing the regional and aggregate impulse responses in my calibrated model to a comparable version without labor migration.[25] The results imply that two key mechanisms—migration accelerator and regional reallocation—play central roles in driving internal propagation effects on regional and aggregate dynamics.

2.6.1 Regional Dynamics and Migration Accelerator Mechanism

2.6.1.1 Regional Impulse Responses

I first show how regional reallocation affects regional dynamics by comparing the impulse responses of regional economic variables in my calibrated model to a comparable model without labor migration. Figure 2.3 plots the regional impulse responses to a 1 percent shock of localized government spending relative to steady state output in a given region. The comparable model generates no response in regional population and regional construction. Since the government spending shock decays exponentially, regional economic activity, such as output, consumption, and wages, declines in a similar exponentially decaying pattern. Most existing models share the same features of the comparable model and cannot generate hump-shaped and lengthy responses as in the data.

In contrast, as labor migration is introduced, the calibrated model can generate sizable amplification effects on the response of regional economic variables. A positive

[25] The benchmark can be seen as a special case of the full model in which net directed labor migration is shut down by setting $\psi = 0$. Hence, there is no labor reallocation across regions and thus population shares and regional housing construction are constant in the model.

localized government spending shock leads to an increase of the regional real wage. According to the optimal migration policy (i.e., equation 2.6), higher regional real wages attract more labor migration towards such booming regions. As more labor gradually migrates into such regions, the regional demand for housing goes up. Because of decreasing returns to scale housing production, regional construction spending increases proportionally more than the increase of population. Thus, regional demand is amplified due to labor migration and regional construction booms. The increasing demand for regional housing construction further drives up regional wages and attracts more labor influx. This leads to a lengthy and hump-shaped response of regional output. I refer this channel as to the "migration accelerator" mechanism through which regional demand shocks can derive amplification on regional economic activity. This mechanism is consistent with recent empirical findings about labor migration and local housing construction (e.g., Howard, 2018).

2.6.1.2 Role of Labor Migration: Calvo and Gumbel Specification

I now explore the quantitative role of Calvo and Gumbel specification in determining the responses of labor migration and regional population. Figure 2.4 shows the regional impulse responses to a 1 percent shock of localized government spending relative to steady state output in a region with 10 percent population at steady state (e.g., California).

As regional demand booms, the optimal migration policy implies that, for an individual agent who is able to relocate, the probability of migrating towards such a region goes up by 12 percent from 0.1 to 0.112 (Panel b). Given the Calvo migration probability at an annual frequency, $\psi = 0.031$, the regional population share evolves according to $\mu_{jt+1} = (1 - \psi)\mu_{jt} + \psi\varphi_{jt}$. Regional population gradually responds to a government spending shock, leading to a 2 percent increase in population after 10

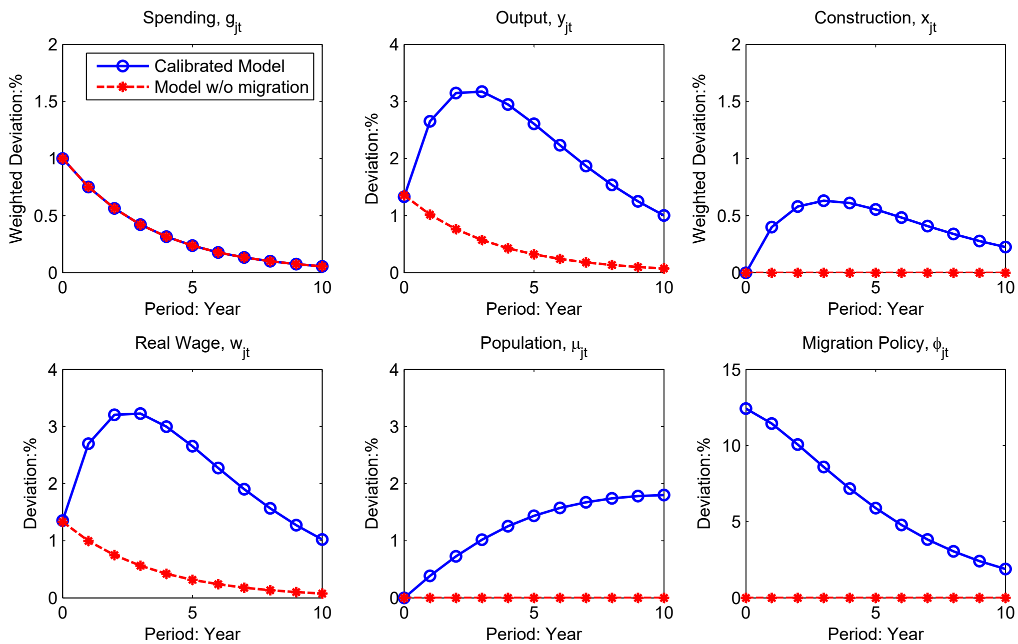


Figure 2.3: Model-based Regional Impulse Responses

Notes: The blue circles and red stars represent the model-based regional impulse responses to 1 percent localized government spending shock relative to output. “Weighted deviation” is the regional impulse responses of a variable weighted by its share in steady state output. The shares of government spending and construction in steady state output are 10 and 5 percent respectively.

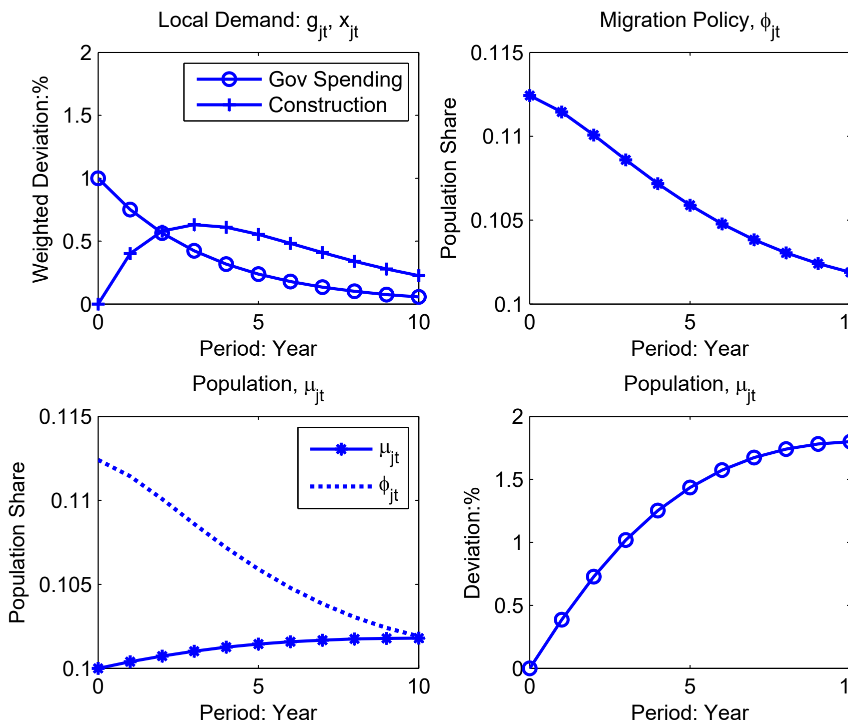


Figure 2.4: Role of Calvo and Gumbel Specification

Notes: Blue lines represent the model-based regional impulse responses to 1 percent localized government spending shock relative to output. "Weighted deviation" is the regional impulse responses of a variable weighted by its share in steady state output. The shares of government spending and construction in steady state output are 10 and 5 percent, respectively.

years relative to the pre-shock level.

This result implies that the Calvo migration probability and Gumbel location-preference shocks play important roles in shaping the gradual response of regional population to a regional demand shock. Intuitively, the Calvo migration probability ψ controls the persistence of the response in regional population, while the Gumbel elasticity ν governs the intensity of the response. In order to generate the *dynamic* responses of regional populations as in the data, one needs to incorporate proper migration frictions, for instance, the Calvo migration probability, to generate reasonable persistence.[26]

[26]Most of models in the spatial economics literature are embedded with Gumbel location-preference shocks but abstract from Calvo migration probability. Under those models, population

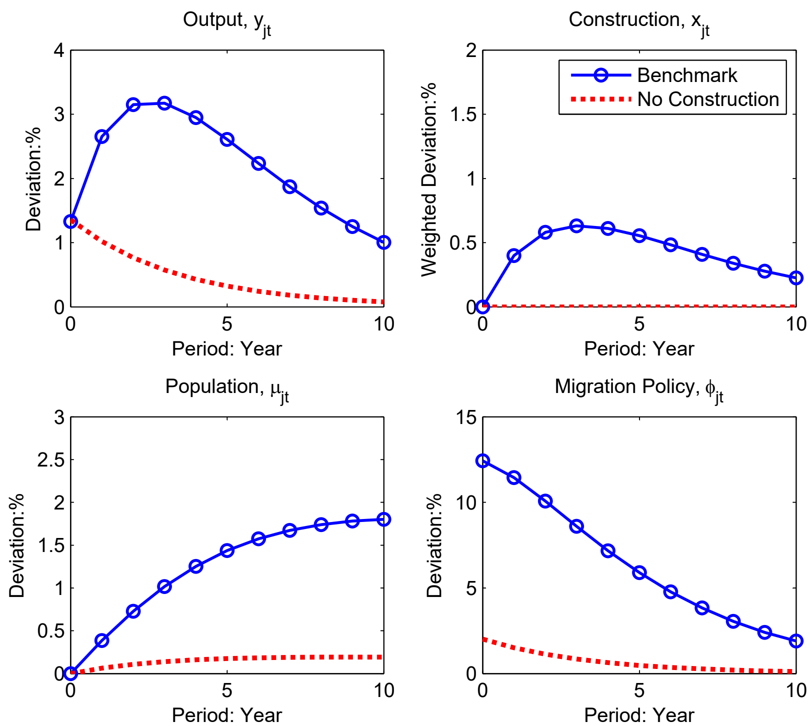


Figure 2.5: Role of Regional Housing Construction

Notes: Blue circles stand for the model-based regional impulse responses to 1 percent localized government spending shock relative to output in the benchmark case, while red dotted lines represent the "no-construction-response" case. "Weighted deviation" is the regional impulse responses of a variable weighted by its share in steady state output.

2.6.1.3 Role of Regional Housing Construction

To highlight the role of regional housing construction, I compare my results to a no-construction-response case. In the no-construction-response case, regional housing construction is a constant and has no response to any changes of labor immigration.

Figure 2.5 reports the comparison. The result shows that, without the response of regional construction to immigration, the regional amplification is substantially weakened. Under the no-response scenario, only a small fraction of labor migration is triggered by localized government spending shocks. Since there is no additional

responds immediately and then converges to the pre-shock level, instead of generating a gradual and persistent pattern.

increase in regional housing construction as a response (dotted line in Panel b), individual agents expect no further prosperity from regional construction demand and thus reduce their propensity to migrate relative to the benchmark case (Panel d). Regional population thus grows at a low speed relative to the case with regional housing construction (Panel c). Therefore, regional output expands substantially less than the case with regional housing construction (Panel a).[27]

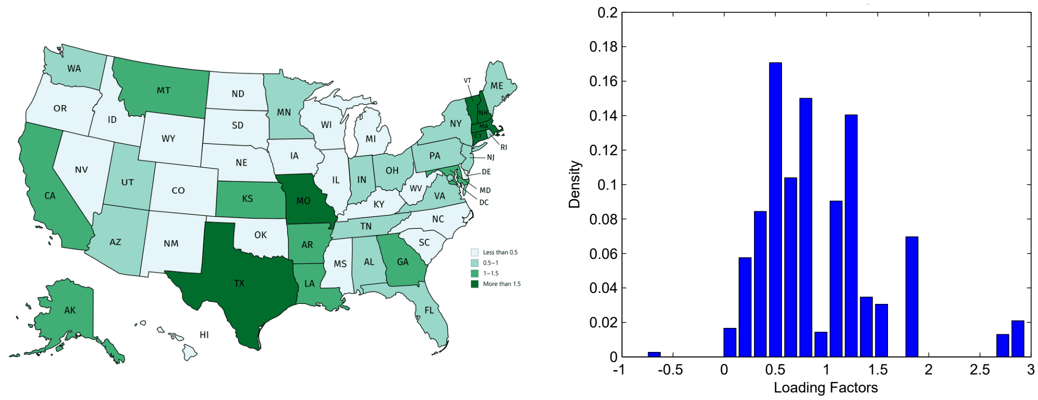
2.6.2 Aggregate Dynamics and Regional Reallocation

Having elucidated the mechanisms and discussed what drives the amplification of regional output, I now quantify the aggregate impacts of asymmetric localized government spending shocks and elaborate the macro consequences of regional reallocation. In addition, I conduct various quantitative experiments and investigate the role of asymmetry and persistence of fiscal policy across regions. .

2.6.2.1 Amplification Effects of Regional Reallocation

Based on the discussion in Chapter 1, state-level military procurement displays systematic heterogeneity in sensitivity to national military spending. For instance, in response to one unit of change in national military spending, the change of military procurement in California is more sensitive to the national one than the change in Illinois is. In fact, the estimated slope coefficients in the first stage regression can be interpreted as the loading factors of these sensitivities. Figure 2.6 displays the distribution of loading factors across states. This result suggests that there is substantial

[27] In case that regional construction is responsive to immigration, the amplification effects on regional economic activity are associated with the permit completion rate, ϕ_h. The smaller completion rate of housing permits leads to a more persistent impact on future construction spending. Therefore, regional construction demand displays a persistent response to labor immigration. Thus, the permit completion rate gives rise to a persistent and hump-shaped response of regional construction to a regional demand shock.



(a) Loading Factors across States (b) Density of Loading Factors

Figure 2.6: Distribution of Loading Factors (Local Sensitivities)

Notes: The loading factors are the estimated slope coefficients in the first-stage regression. The left panel shows the loading factors across U.S. states in a map. The right panel reports the density of loading factors weighted by average population shares across states (where Mississippi has a negative sensitivity).

heterogeneity of state military spending in sensitivity to national military spending.

Using the loading factors across regions as heterogeneous regional sensitivities, I quantify the aggregate impacts of 1 percent national military buildup relative to output that affects state military spending differentially in the calibrated model. Figure 2.7 displays the aggregate impulse responses. In response to an increase in government spending, aggregate consumption drops due to crowding-out effect (Panel c). Therefore, aggregate labor supply increases (Panel f) as well as more capital investment is undertaken (Panel e). Hence, aggregate output increases. This is consistent with common wisdom.

In contrast, the asymmetric localized government spending shocks give rise to labor reallocation across regions. As shown in the above discussion, the regional shocks trigger labor migration towards booming regions. As more labor migrates into those regions, regional housing demand goes up, driving an increase of construction spending in a greater proportion than the increase of the regional population. The increas-

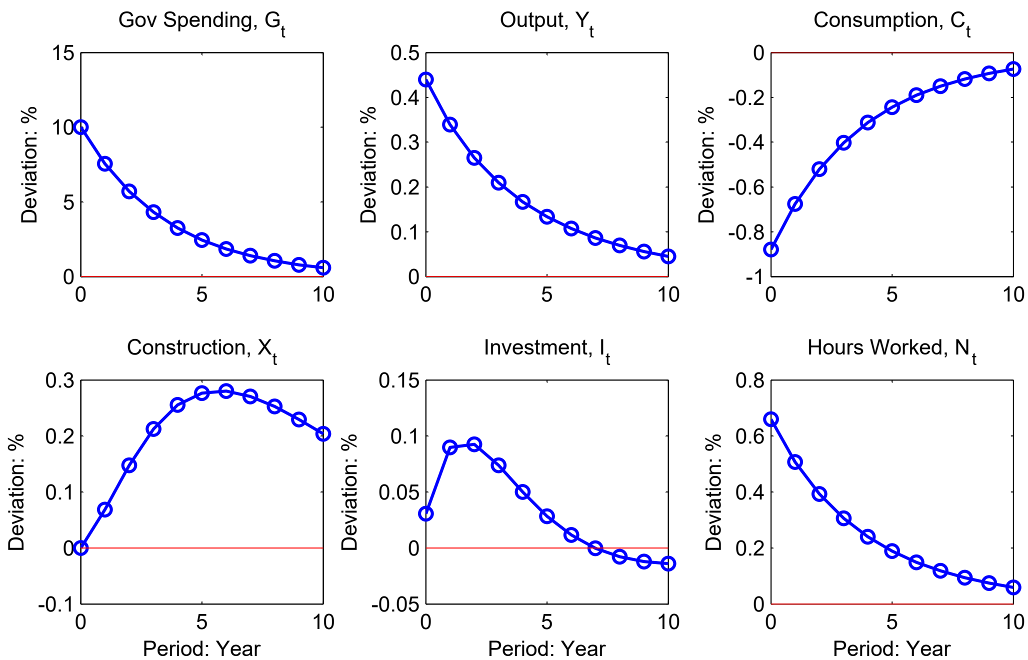


Figure 2.7: Aggregate Impulse Responses

Notes: Blue lines represent the model-based aggregate impulse responses to 1 percent national military spending shock relative to steady state output that affects states differently according to the estimated loading factors as local sensitivities shown in Figure 2.6. Consumption/output ratio is 65 percent, investment/output ratio is 20 percent, government-purchase/output ratio is 10 percent, and construction-spending/output ratio is 5%.

ing construction demand triggers further increase of regional wages, attracting even more labor over time. The labor migration accelerator amplifies the magnitude of regional demand changes. In addition, given increasing regional construction spending per resident in booming regions, a net increase of aggregate construction spending results, simply because net directed migration reallocates more population to booming regions, which get more weights in aggregate. Therefore, aggregate construction rises in response to labor reallocation across regions and serves as an aggregate demand shifter, which creates a positive covariance effect: the regional reallocation mechanism. This force amplifies the response of aggregate output, eventually leading to a larger aggregate fiscal multiplier.

To evaluate the overall impacts of fiscal stimulus over various horizons, I calculate the cumulative aggregate fiscal multipliers to output as defined in Ilzetzki, Mendoza, and Vegh (2013).[28] To show the impact of the regional reallocation channel, I compute the cumulative aggregate fiscal multiplier in the case of shutting down the possibility of labor migration (i.e., $\psi = 0$). Figure 2.8 reports the results with and without labor migration. Quantitatively, labor reallocation leads to a 30 percent amplification of the cumulative aggregate output multiplier relative to a benchmark model without labor migration (0.56 vs. 0.43), given the spatial heterogeneity of regional military spending shocks.

[28] They define the cumulative multiplier at time T by

$$\text{Cumulative Multiplier}(T) = \frac{\sum_{t=0}^{T}(1+i)^{-t}\Delta y_t}{\sum_{t=0}^{T}(1+i)^{-t}\Delta g_t},$$

where i is the steady state interest rate and Δy_t and Δg_t represent the changes of aggregate output and government spending between period t and $t-1$.

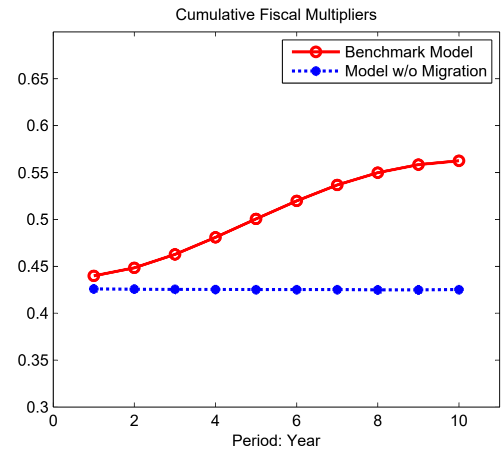


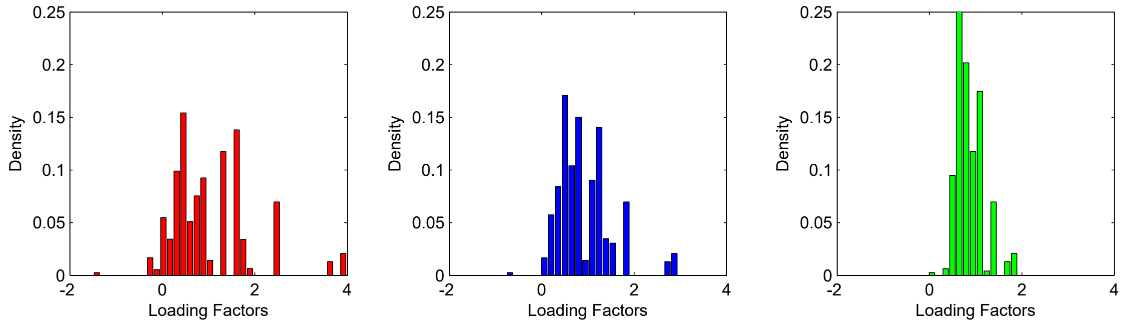
Figure 2.8: Cumulative Aggregate Fiscal Multiplier

Notes: The cumulative aggregate fiscal multipliers are based on the impulse responses of aggregate output to a 1 percent national military spending shock relative to steady state output that affects states differently as in Figure 2.6. Red circles represent the multipliers in the benchmark calibrated model, and blue stars stand for the model without labor migration by setting $\psi = 0$.

2.6.2.2 The Impacts of Asymmetry

I analyze how the aggregate effects of national fiscal policies depend on the cross-sectional distribution of fiscal stimulus. I apply the distribution of loading factors in Figure 2.6 as a benchmark and quantify the role of regional heterogeneity by altering the distribution of loading factors. Applying a mean-preserving spread, I stress the magnitudes of loading factors and generate two alternative distributions of localized government spending shocks as being 50 percent more dispersed (Case 1) and 50 percent less dispersed (Case 2) relative to the benchmark case. Figure 2.9 displays the alternative distribution of loading factors. In response to 1 percent national government spending shock relative to steady state output, the localized government spending shocks are more asymmetric in Case 1 than in the benchmark case, while the shocks are less asymmetric in Case 2 than in the benchmark case.[29]

[29] In terms of cross-sectional standard deviation of the distribution across states, the benchmark case has a standard deviation of 0.63, Case 1 of 0.95, and Case 2 of 0.32.



(a) 50% more dispersed (b) Benchmark (c) 50% less dispersed

Figure 2.9: Alternative Distribution of Loading Factors

Notes: The figure reports the density of loading factors weighted by average population shares across states. The loading factors are the estimated slope coefficients in the first-stage regression.

Figure 2.10 shows the cumulative aggregate output multipliers for the benchmark and alternative cases. These results suggest that the aggregate impact of regional demand shocks crucially depends on their cross-sectional dispersion. If regional demand shocks are more asymmetric, labor reallocation triggers a larger magnitude of net directed migration. It gives rise to a larger amount of regional construction spending as regional housing demand increases in booming regions. Aggregating the changes of regional demand, a larger increase of aggregate demand follows due to a larger covariance effect. Thus, aggregate output increases even more, leading to a larger cumulative fiscal multiplier. In Case 1 with more asymmetric distribution of regional sensitivities, the aggregate fiscal multiplier is amplified by 46 percent relative to a comparable model without labor migration (0.63 vs. 0.43); while in Case 2 with a 50 percent less asymmetric distribution, the amplification drops to 12 percent (0.48 vs 0.43). If all states suffer from the same regional demand shocks, no labor reallocation is triggered. Thus, the regional reallocation channel is shut down and the aggregate impacts are the same as the model without labor reallocation predicts. The degree of asymmetry of fiscal policy across regions is quantitatively important

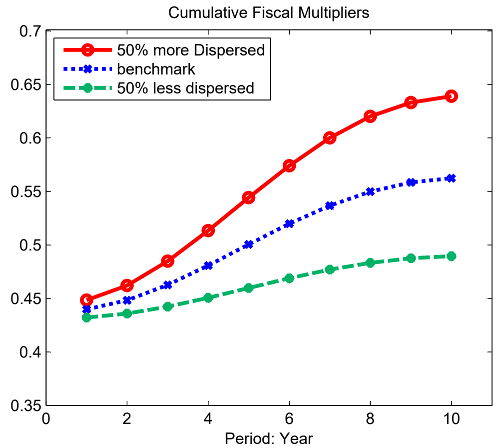


Figure 2.10: Aggregate Cumulative Multipliers: Distributional Dependence
Notes: The figure reports the cumulative aggregate output multipliers in response to a 1 percent increase in government spending relative to output that affects states differently according to alternative distributions of loading factors: 50% more dispersed, benchmark, and 50% less dispersed.

for the evaluation of aggregate fiscal multipliers.

2.6.2.3 The Impacts of Persistence

I now investigate how the amplification effect of regional reallocation on aggregate output also depends on the persistence of fiscal policy. Intuitively, even if a fiscal shock affects regions differentially, individuals have a smaller migration propensity in response to a transitory fiscal shock relative to a persistent shock. This also affects the magnitude of the amplification effect of regional reallocation.

To explore the sensitivity to persistence, I consider a case with relatively low persistence of government spending shocks. I specify the persistence at an annual frequency to be $\rho_z = 0.56$, where the half-life of the shock is roughly one year and the impact of the shock essentially returns to a normal level after three years; in the benchmark case, the persistence is 0.75 and the impact of the shock mean-reverts to a normal level after six years.

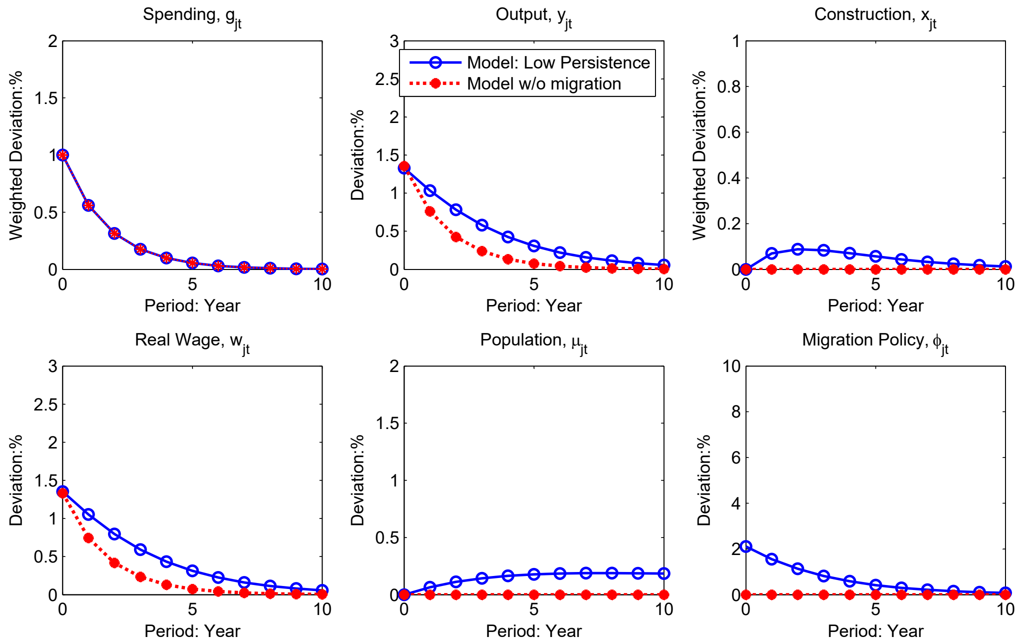


Figure 2.11: Regional Impulse Responses: Low Persistence

Notes: Blue circles and black stars represent the model-based regional impulse responses to a 1 percent government spending shock relative to output with low persistence, $\rho_z = 0.56$. "Weighted deviation" is the regional impulse responses of a variable weighted by its share in output. The shares of government spending and construction in output are 10 and 5 percent respectively.

Figure 2.11 plots the regional impulse responses to a 1 percent localized government spending shock relative to steady state output with relatively low persistence $\rho_z = 0.56$. The magnitude of labor immigration is substantially weakened, compared to the benchmark model, and leads to a moderate increase in regional housing demand and regional construction. As Panel (b) shows, the amplification effect on regional output per resident is significantly smaller than in the benchmark case.

Figure 2.12 shows the cumulative aggregate fiscal multipliers in response to a relatively low persistent shock for the models with and without labor migration. The amplification effect of regional reallocation on aggregate output is only 5 percent, where the ten-year cumulative aggregate fiscal multiplier is 0.32 for the case with

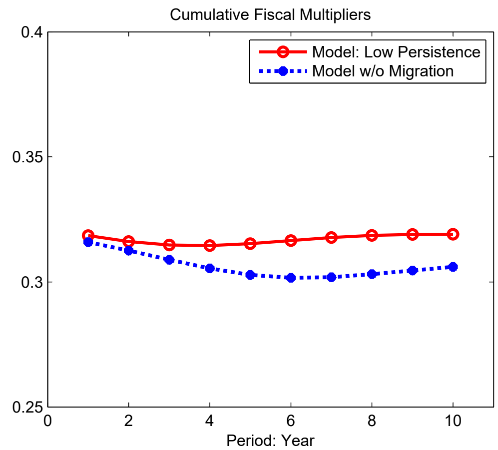


Figure 2.12: Cumulative Aggregate Fiscal Multiplier: Low Persistence

Notes: The cumulative multipliers are based on the impulse responses of aggregate output to a 1 percent national military spending shock relative to output with low persistence of $\rho_z = 0.56$ that affects states differently as in Figure 2.6. Red circles represent the multipliers in the model with labor migration, and blue stars stand for the model without labor migration by setting $\psi = 0$.

labor migration and is 0.305 for the case with fixed population. This is because individuals have a smaller migration propensity in response to a transitory shock and therefore there is less of a stimulating effect on regional output per resident through a regional construction boom. This leads to a relatively smaller covariance effect and thus mitigates the amplification effect of regional reallocation on aggregate output.

2.7 Conclusion

This chapter investigates the macroeconomic implications of regional reallocation in affecting both regional and aggregate fiscal multipliers. Exploiting state-level variation of military spending, I provide evidence that the regional consequences are large and lengthy because the asymmetric shocks trigger regional reallocation of population. Labor reallocation amplifies regional output through a boom in construction spending and aggregate output through a covariance effect arising from net directed migration towards booming regions. Using a new method that is both tractable and sufficiently rich to capture heterogeneity across regions, my analysis shows that regional reallocation and systematic regional heterogeneity are quantitatively very important for aggregate output effects of fiscal policy. Virtually every macroeconomic shock affects regions differently, so that many shocks may lead to the regional reallocation effect similar to this paper. Given the pervasive nature of heterogeneity in regions and macroeconomic shocks, future work should analyze the importance of regional reallocation for monetary policy, trade shocks, productivity shocks, etc.

Appendix

A Theoretical Derivations

A.1 Proofs of Propositions in Section 2.2

Proposition 1. *Under spatial equilibrium, the exact solution for regional output per resident in response to government purchase shocks is given by*

$$y_j = y^\star + an_g \cdot (g_j - g^\star) + an_x \cdot (x_j - x^\star)$$

and the exact solution for aggregate output is

$$Y = Y^\star + \underbrace{an_g \cdot (\sum_j G_j - G^\star)}_{\text{mean effect}} + \underbrace{an_x \cdot \sum_j (x_j - x^\star)(\mu_j - \mu^\star)}_{\text{covariance effect}},$$

where regional per resident government purchases is $g_j = G_j/\mu_j$, construction spending is $x_j = \mu_j/L$, and equilibrium regional population is $\mu_j = \mu(G_j, \{G_l\}_{l\in\mathcal{J}})$.

Proof: Assume that G_j are random shocks and asymmetric across regions. Given a quadratic utility function $u(c,n) = \frac{1}{2}(\bar{c}-c)^2 - \frac{\lambda_n}{2}n^2$, inelastic demand with one unit of house for each individual $h_j = 1$, and a Cobb-Douglas housing production function $h_j = (\frac{L}{\mu_j})^{\frac{1}{2}}(x_j)^{\frac{1}{2}}$, the equilibrium conditions with regional reallocation are given by

$$\begin{aligned}
\lambda_n n_j &= a(\bar{c} - c_j),\\
a n_j &= c_j + x_j + G_j/\mu_j,\\
x_j &= \mu_j/L,\\
\mathcal{U} &= u(c_j, n_j),\\
1 &= \Sigma_j \mu_j,
\end{aligned}$$

where $\mathcal{U}$ denotes the equalized utility across regions in a spatial equilibrium. Thus, regional consumption and labor supply are equalized across regions in equilibrium. High spending regions attract labor immigration such that regional population μ_j is determined by the equalization of $G_j/\mu_j + \mu_j/L$ across regions. The equilibrium distribution of population is given by $\mu_j = \mu(G_j, \{G_l\}_{l\in\mathcal{J}})$, the equilibrium regional labor supply function is given by

$$n_j = \frac{\bar{c} + G + X}{a + \lambda_n/a},$$

where $X = \sum_j x_j \mu_j$ is total construction spending, $x_j = \mu_j/L$ is regional construction spending, and $\mu_j = \mu(G_j, \{G_l\}_{l\in\mathcal{J}})$ is equilibrium distribution of population. The equilibrium distribution of population across regions is determined by the following conditions: $G_j/\mu_j + \mu_j/L = m$ equalizes for all j, and $\sum_j \mu_j = 1$.

Therefore, the aggregate output is given by $Y = an = \frac{\bar{c}+G+X}{1+\lambda_n/a^2}$. Since the aggregate output is linear in G and X, the first-order expansion equals to the exact solution,

$$Y = Y^{\star} + a^{\star} n_g \cdot (G - G^{\star}) + a^{\star} n_x \cdot (X - X^{\star}),$$

where $n_g = (a + \lambda_n/a)^{-1}$ and $n_x = (a + \lambda_n/a)^{-1}$ are the first-order derivatives with respect to G and X. Using a covariance decomposition identity, the aggregate construction spending X can be represented by

$$X = X^{\star} + \sum_j x^{\star}(\mu_j - \mu^{\star}) + \sum_j (x_j - x^{\star})\mu_j^{\star} + \sum_j (x_j - x^{\star})(\mu_j - \mu^{\star}),$$

Due to $\sum_j \mu_j = \sum_j \mu^{\star} = 1$ and $x_j = \mu_j/L$, the first and second terms are equal to zero, $\sum_j x^{\star}(\mu_j - \mu^{\star}) = 0$ and $\sum_j (x_j - x^{\star})\mu^{\star} = 0$. Hence, the aggregate output Y can be represented by

$$Y = Y^{\star} + \underbrace{a^{\star} n_G \cdot (\sum_j G_j - G^{\star})}_{\text{mean effect}} + \underbrace{a^{\star} n_X \cdot \sum_j (x_j - x^{\star})(\mu_j - \mu^{\star})}_{\text{covariance effect}}, \tag{2.7}$$

where $x_j = \mu_j/L$ and $\mu_j = \mu(G_j, \{G_l\}_{l \in \mathcal{J}})$ indicating that high spending regions attract more labor immigration.

Q.E.D.

Proposition 2. *Under spatial equilibrium, the first-order approximation of the response of regional output to government purchase shocks is given by*

$$y_j = y^{\star} + a n_g \cdot (g_j - g^{\star}) + a n_x (x_j - x^{\star}) + h.o.t$$

and the second-order approximation to aggregate output is given by

$$Y = Y^\star + an_g(\sum_j G_j - G^\star) + \underbrace{an_x \sum_j \tilde{x}_\mu(\mu_j - \mu^\star)^2}_{\textit{covariance effect}} + \frac{1}{2}an_{gg}(\sum_j G_j - G^\star)^2 + h.o.t.,$$

where $\tilde{x}_\mu \equiv (x_\mu + \frac{1}{2}x_{\mu\mu}\mu^\star) > 0$*,* x_μ *and* $x_{\mu\mu}$ *are first- and second-order derivatives of construction spending with respect to regional population.*

Proof: Assume that G_j are random shocks and asymmetric across regions. Therefore, the equilibrium conditions are given by

$$\begin{aligned} v_n(n_j) &= a \cdot u_c(c_j), \\ an_j &= c_j + x_j + G_j/\mu_j, \\ 1 &= h(L/\mu_j, x_j), \\ \mathcal{U} &= u(c_j) - v(n_j), \\ 1 &= \Sigma_j \mu_j, \end{aligned}$$

where $\mathcal{U}$ denotes the equalized utility across regions in a spatial equilibrium. Thus, under spatial equilibrium, regional labor supply per resident is equalized across regions as well as consumption per resident. It also implies that $m \equiv x_j + G_j/\mu_j$ equalizes across regions. Thus, the equilibrium labor supply function is $n_j = n(m)$ such that $an_j = c(n_j) + m$. The regional output per resident is given by $y_j = a \cdot n(x_j + g_j)$. The first order expansion of regional output per resident around steady state can be shown by

$$y_j = y^\star + an_g \cdot (g_j - g^\star) + an_x \cdot (x_j - x^\star) + h.o.t.,$$

where n_g and n_x are the first-order derivatives of regional labor supply per resident with respect to government purchases and construction spending per resident, and $h.o.t.$ denotes higher order terms. Knowing that u_c is a decreasing function in c and v_n is an increasing function in n, this yields c is a decreasing function in n in equilibrium. Therefore, $n_g > 0$ and $n_x > 0$ due to the implicit function theorem.

Given $m = \sum_j x_j \mu_j + \sum_j G_j = \sum_j x(\mu_j)\mu_j + G$, the aggregate output can be represented by $Y = a \cdot n(X + G)$. The second-order expansion of aggregate output can be shown by $Y = Y^\star + an_g(G - G^\star) + an_x(X - X^\star) + \frac{1}{2}an_{gg}(G - G^\star)^2 + \frac{1}{2}an_{xx}(X - X^\star)^2 + h.o.t.$, where n_{gg} and n_{xx} are the second-order derivatives of regional labor supply per resident with respect to government purchases and construction spending per resident. The second-order expansion of aggregate construction spending $X = \sum_j x(\mu_j)\mu_j$ is given by $X = X^\star + \sum_j x_\mu(\mu_j - \mu^\star)\mu^\star + \sum_j x^\star(\mu_j - \mu^\star) + \sum_j x_\mu(\mu_j - \mu^\star)^2 + \frac{1}{2}\sum_j x_{\mu\mu}\mu^\star(\mu_j - \mu^\star)^2 + h.o.t.$ Due to the adding up condition $\sum_j \mu_j = \sum_j \mu^\star = 1$, the first-order terms wash out. Hence, the aggregate construction spending is given by

$$X = X^\star + \sum_j \tilde{x}_\mu(\mu_j - \mu^\star)^2 + h.o.t.,$$

where $\tilde{x}_\mu \equiv x_\mu + \frac{1}{2}x_{\mu\mu}\mu^\star$, x_μ and $x_{\mu\mu}$ are first and second order derivatives of construction spending with respect to regional population. Therefore, the Taylor expansion of the aggregate output can be represented by

$$\begin{aligned} Y = Y^\star + an_g(\sum_j G_j - G^\star) + an_x \sum_j \tilde{x}_\mu(\mu_j - \mu^\star)^2 + \frac{1}{2}an_{gg}(\sum_j G_j - G^\star)^2 \\ + \frac{1}{2}an_{xx}\sum_j \sum_l \tilde{x}_\mu^2(\mu_j - \mu^\star)^2(\mu_l - \mu^\star)^2 + h.o.t., \end{aligned}$$

According to equilibrium population distribution $\mu_j = \mu(G_j, \{G_l\})$, the expansion of regional population is given by

$$\mu_j = \mu^\star + \mu_{1G}(G_j - G^\star) + \sum_l \mu_{2G}(G_l - G^\star) + h.o.t.,$$

where μ_{1G} and μ_{2G} are the first-order derivatives of μ_j with respect to G_j and G_l respectively. Therefore, $\sum_j(\mu_j - \mu^\star)^2$ corresponds to a second-order term $\mu_{1G}^2 \sum_j (G_j - G^\star)^2 + (J\mu_{2G}^2 + \mu_{1G}\mu_{2G}) \cdot (G - G^\star)^2$, while $\sum_j \sum_l (\mu_j - \mu^\star)^2(\mu_l - \mu^\star)^2$ corresponds to a fourth-order term. As a result, the second-order expansion of the aggregate output is given by

$$Y = Y^\star + an_g(\sum_j G_j - G^\star) + an_x \sum_j \tilde{x}_\mu(\mu_j - \mu^\star)^2 + \frac{1}{2}an_{gg}(\sum_j G_j - G^\star)^2 + h.o.t.$$

Alternatively, in terms of regional government spending shocks, we have

$$Y = Y^\star + an_g(\sum_j G_j - G^\star) + an_x\tilde{x}_\mu\mu_{1G}^2 \sum_j (G_j - G^\star)^2 + an_x\tilde{x}_\mu\tilde{\mu}_G(\sum_j G_j - G^\star)^2$$
$$+ \frac{1}{2}an_{gg}(\sum_j G_j - G^\star)^2 + h.o.t.$$

where $\tilde{\mu}_G \equiv J\mu_{2G}^2 + \mu_{1G}\mu_{2G}$.

Q.E.D.

A.2 The Equilibrium of Regional Housing Market

In the benchmark model in Section 4, an individual household in region j chooses consumption and housing services at relative rental prices ϱ_{jt}^h. Regional population evolves according to $\mu_{j,t+1} = (1-\psi)\mu_{jt} + \psi\varphi_{jt}$, where $\psi\varphi_{jt}$ denotes the amount of new immigrants. In each period a fraction δ_h of housing stock depreciates such that the local housing stock evolves as $H_{jt+1} = (1-\delta_h)H_{jt} + Y_{jt}^h$, where Y_{jt}^h denotes the amount of houses built at time t. Assume that there is a continuum of competitive builders whose size is the same as local population. Each builder operates a decreasing return to scale housing production technology, $y^h_{jt} = bx^{\zeta}_{jt}$, where $\zeta \in (0, 1)$ and x_{jt} and b are the construction input and the productivity of the housing sector. Thus, the total amount of newly-built houses is $Y^h_{jt} = y^h_{jt}\mu_{jt}$.

In order to generate hump-shaped responses of regional construction, the model needs to introduce some frictions on regional housing supply. Permit restrictions are a natural friction. In practice, the government issues a limited number of permits to restrict the regional supply of housing due to political, environmental, and other reasons. In this model, I assume that the government issues one unit of housing permit for each new immigrant.[30] I also assume that a fraction ϕ_h of the authorized permit stock S_{jt} is completed in each period, where $\phi_h \in (0, 1)$ denotes the completion rate. Thus, new houses built up at time t is not greater than the amount of completed permits, $Y^h_{jt} \leq \phi_h S_{jt}$. The stock of authorized permits S_{jt} is the sum of the unfinished permits for the last period $(1 - \phi_h)S_{jt-1}$ and new issued permits $S^n_{jt} = \psi\varphi_{jt}$, given by $S_{jt} = (1 - \phi_h)S_{jt-1} + \psi\varphi_{jt}$.

Therefore, builders optimally choose the amount of houses to build subject to permit restrictions, max : $q_{jt}y^h_{jt} - x_{jt}$, s.t. $y^h_{jt} = bx^{\zeta}_{jt}$ and $y^h_{jt}\mu_{jt} \leq \phi_h S_{jt}$. The equilibrium rental rate and housing prices are pinned down by the individual household's marginal rate of substitution between consumption and housing and the user cost equation.

The equilibrium conditions in regional housing markets are given by (i) the optimal

[30] Howard and Liebersohn (2018) establish an empirical fact that the change of population is nearly one-for-one with the change of housing units, implying that labor reallocation is the main force for cross-sectional housing demand. The "one permit per immigrant" assumption is consistent with this empirical fact.

condition for individual's housing demand is

$$u_c(c_{jt}) \cdot \varrho_{jt} = u_h(h_{jt}), \tag{2.8}$$

(ii) per capita housing services is

$$h_{jt} = H_{jt}/\mu_{jt},$$

(iii) the resource constraint is

$$c_{jt} + x_{jt} + g_{jt} = y_{jt},$$

(iv) housing construction restricted by completed permit is

$$\phi_h S_{jt} = b x_{jt}^{\zeta} \mu_{jt}, \tag{2.9}$$

$$q_{jt} \zeta b x_{jt}^{\zeta - 1} > 1, \tag{2.10}$$

(v) the user cost condition for housing prices and rents is

$$q_{jt} = \beta \mathbb{E}_t \Lambda_{t+1} \left(\varrho_{jt+1} + (1 - \delta_h) q_{jt+1} \right), \tag{2.11}$$

(vi) the evolution of local housing stock is

$$H_{jt+1} = (1 - \delta_h) H_{jt} + \phi_h S_{jt},$$

(vii) the evolution of authorized permit stock is

$$S_{jt+1} = (1 - \phi_h) S_{jt} + \psi \varphi_{jt},$$

and (viii) the evolution of local population is

$$\mu_{jt+1} = (1-\psi)\mu_{jt} + \psi\varphi_{jt}.$$

Given regional output y_{jt}, regional government spending g_{jt} and new immigrants $\psi\varphi_{jt}$, the above equilibrium conditions pin down $\{c_{jt}, h_{jt}, \varrho_{jt}, x_{jt}, q_{jt}, H_{jt}, S_{jt}, \mu_{jt}\}$.

Now focus on builders optimality conditions and check whether the permit restriction binds at (and around) steady state. Specify the utility function be $u(c,h) = \log c + \chi_h \log h$. The equilibrium conditions at steady state are given by

$$\begin{aligned}
\varrho_j^\star &= \chi_h c_j^\star / h_j^\star, \\
h_j^\star &= H_j^\star / \mu_j^\star, \\
y_j^\star &= c_j^\star + x_j^\star + g_j^\star, \\
\phi_h S_j^\star &= \psi \mu_j^\star, \\
\psi \mu_j^\star &= b x_j^{\star\zeta} \mu_j^\star, \\
\delta_h H_j^\star &= \psi \mu_j^\star, \\
q_j^\star &= \frac{\beta \varrho_j^\star}{1-\beta(1-\delta_h)}, \\
q_j^\star \zeta b x_j^{\star\zeta-1} &> 1.
\end{aligned}$$

Thus, substitute all the equality conditions and the calibrated values in Section 6, we have

$$\begin{aligned}
q_j^\star \zeta b x_j^{\star\zeta-1} &= \frac{\beta\zeta}{1-\beta(1-\delta_h)} \frac{\varrho_j^\star b x_j^{\star\zeta}}{x_j^\star} = \frac{\beta\zeta\delta_h\chi_h}{1-\beta(1-\delta_h)} \frac{c_j^\star}{x_j^\star} \\
&= \frac{0.97 \times 0.74 \times 0.03 \times 0.22}{1-0.97\times(1-0.03)} \times \frac{0.70}{0.05} = 1.12 > 1.
\end{aligned}$$

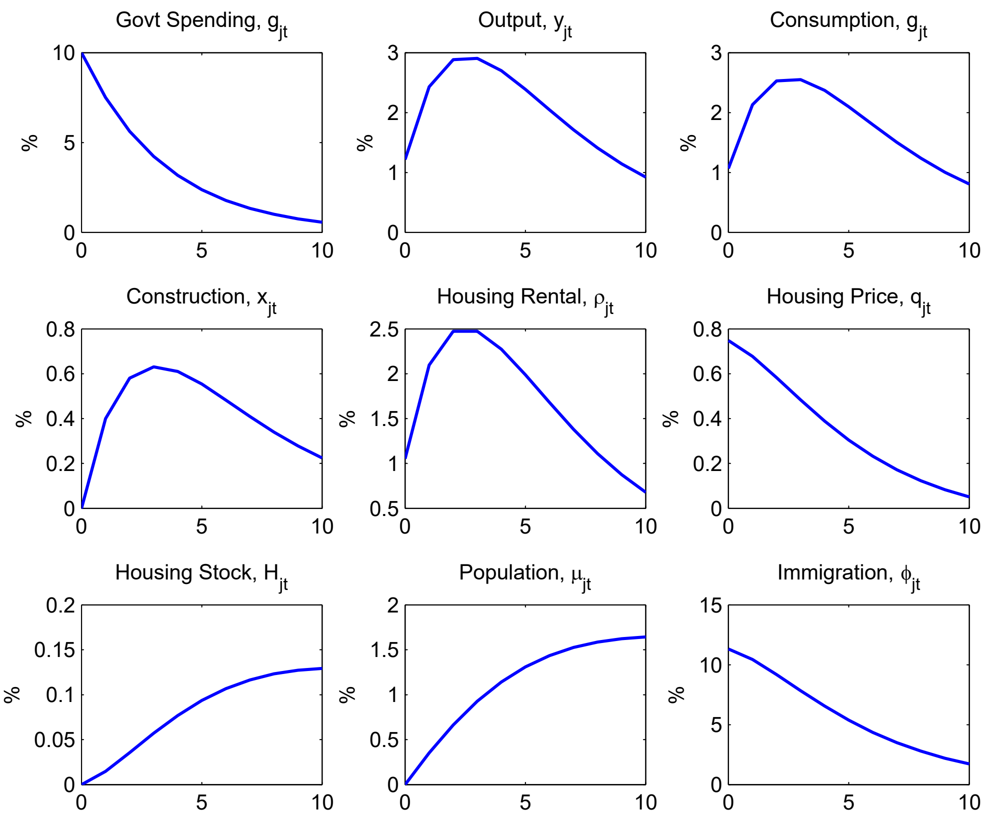


Figure 2.13: Regional Impulse Responses to a Localized Government Spending Shock

Notes: The blue curves represent the model-based regional impulse responses to 1 percent localized government spending shock relative to steady state output, in the benchmark model with migration frictions and regional housing markets.

Under the benchmark calibrated values of parameters, the builders would like to build more houses than the amount of authorized permits issued by the government under "one person one permit" assumption.

Figure 2.13 shows the regional impulse responses to a localized government spending shock based on the benchmark model with migration frictions and housing permit restrictions. As more new immigrants move to the region, the housing stock increases to maintain individual's housing consumption but not immediately. Under the limited completion specification, the regional housing stock increases less than the regional population does in the short run. Therefore, per resident housing consumption drops and housing rental rate rises, which leads to an increase in regional housing prices.

A.3 The Optimal Migration Policy

This subsection shows the derivation of optimal migration policy and Bellman equation.

A.3.1 The Optimal Migration Policy

Lemma 1. Given the household's optimization problem in the benchmark model,

$$V_j(\mathbf{s}_t) = \max_{\{c_{jt}, n_{jt}, h_{jt}, \varphi_t^\star\}} : \left\{ u(c_{jt}, n_{jt}, h_{jt}) + \kappa_j + \beta \left[\begin{array}{c} \psi\left(\max_{l \in \mathcal{J}} : \mathbb{E}_t[V_l(\mathbf{s}_{t+1}) + \nu\epsilon_{lt}]\right) \\ +(1-\psi)\mathbb{E}_t V_j(\mathbf{s}_{t+1}) \end{array} \right] \right\}$$

subject to budget constraint, $c_{jt} + \varrho_{jt}^h h_{jt} = (1-\tau_t)(W_{jt}/P_{jt})n_{jt} + D_t + R_t^m - T_t$. Under Gumbel distribution assumption upon ϵ_{lt}, the optimal migration policy is represented by

$$\varphi_{jt}^\star = \frac{\exp\left[\mathbb{E}_t V_j(\mathbf{s}_{t+1})\right]^{\frac{1}{\nu}}}{\sum_l \exp\left[\mathbb{E}_t V_l(\mathbf{s}_{t+1})\right]^{\frac{1}{\nu}}}.$$

Proof: (i) Denote $\tilde{V}_{jt} \equiv E[\mathbb{E}_t V_{jt+1} + \nu\epsilon_{jt}]$ by the expected future utility of an individual household locating in region j at period $t+1$, where the expectation E is taken over the location preference shocks. Assume that the location preference shocks ϵ_{lt} is i.i.d. across individuals and over time, and is a realization of a Gumbel distribution with zero mean. In particular, $F(\epsilon) = \exp[-\exp(-\epsilon - \overline{\gamma})]$, and therefore $f(\epsilon) = \partial F/\partial\epsilon$, where $\overline{\gamma} = \int_{-\infty}^{+\infty} x \exp(-x - \exp(-x))dx$ to Euler's constant. Then denote the expected future value for an individual household who are able to relocate in period t by

$$\Phi_t = E\left[\max_{l \in \mathcal{J}} : \mathbb{E}_t V_{lt+1} + \nu\epsilon_{lt}\right].$$

The optimal migration policy is the solution to achieve the above expected future value.

(ii) Define $\lambda_t \equiv \log \sum_j \exp\{-\frac{1}{v}(\mathbb{E}_t V_{jt+1} - \mathbb{E}_t V_{lt+1})\}$ and consider the following change of variables, $\xi_{jt} = \epsilon_{jt} + \overline{\gamma} - \lambda_t$. I now derive the optimal relocation policy. Define $\varphi^{\star}_{jt}$ as the fraction of those who have relocation opportunity that relocate to region j (no matter where they are in period t). This fraction is equal to the probability that a given individual of relocation opportunity choose to locate in region j at the end of period t; that is to say, the probability that the expected utility of moving to j is higher than the expected utility in any other regions. Formally, this yields

$$\varphi^{\star}_{jt} = \Pr\left(\frac{1}{\nu}\mathbb{E}_t V_{jt+1} + \epsilon_{jt} \geq \max_{l \neq j} : \{\frac{1}{\nu}\mathbb{E}_t V_{lt+1} + \epsilon_{lt}\}\right).$$

Given the Gumbel distribution assumption on the idiosyncratic location preference shocks,

$$\varphi^{\star}_{jt} = \int_{-\infty}^{+\infty} f(\epsilon_{jt}) \Pi_{l \neq j} F(\mathbb{E}_t V_{jt+1} - \mathbb{E}_t V_{lt+1} + \epsilon_{jt}) d\epsilon_{jt}.$$

Substituting for $F(\epsilon)$ and $f(\epsilon)$, we have

$$\varphi^{\star}_{jt} = \int_{-\infty}^{+\infty} \exp(-\epsilon_{jt} - \overline{\gamma}) \exp\left[-\exp(-\epsilon_{jt} - \overline{\gamma}) \sum_l \exp(-\frac{1}{\nu}(\mathbb{E}_t V_{jt+1} - \mathbb{E}_t V_{lt+1}))\right] d\epsilon_{jt}.$$

Using the change of variable $\xi_{jt} = \epsilon_{jt} + \overline{\gamma} - \lambda_t$, this yields

$$\varphi^{\star}_{jt} = \exp(-\lambda_t) \int_{-\infty}^{+\infty} \exp[-\xi_{jt} - \exp(-\xi_{jt})] d\xi_{jt}$$

and solving for this integral, we obtain the optimal migration policy,

$$\varphi^{\star}_{jt} = \frac{\exp[\mathbb{E}_t V_j(\boldsymbol{s}_{t+1})]^{\frac{1}{\nu}}}{\sum_l \exp[\mathbb{E}_t V_l(\boldsymbol{s}_{t+1})]^{\frac{1}{\nu}}}.$$

Q.E.D.

A.3.2 The Bellman Equation

Lemma 2. The Bellman equation of individual's value function is

$$V_j(\mathbf{s}_t) = \max_{\{c_{jt}, n_{jt}, h_{jt}, \varphi_t^\star\}} : \left\{ u(c_{jt}, n_{jt}, h_{jt}) + \kappa_j + \beta \begin{bmatrix} \psi \left(\max_{l \in \mathcal{J}} : \mathbb{E}_t[V_l(\mathbf{s}_{t+1}) + \nu \epsilon_{lt}] \right) \\ +(1-\psi)\mathbb{E}_t V_j(\mathbf{s}_{t+1}) \end{bmatrix} \right\}$$

Under Gumbel distribution assumption upon $\{\epsilon_{lt}\}$, the above Bellman equation can be represented by

$$V_{jt} = u(c_{jt}^\star, n_{jt}^\star, h_{jt}^\star) + \kappa_j + \beta\psi\nu \left(\log \sum_l \exp\left(\frac{1}{\nu} \mathbb{E}_t V_{lt+1} \right) \right) + \beta(1-\psi)\mathbb{E}_t V_{jt+1}.$$

Proof: Given the definition of expected future value for an individual household who is able to relocate in period t by

$$\Phi_t = E\left[\max_{l \in \mathcal{J}} : \mathbb{E}_t V_{lt+1} + \nu \epsilon_{lt} \right],$$

where the expectation operator E is associated with the idiosyncratic location preference shock. According to the Gumbel distribution assumption on idiosyncratic location preference shocks, this yields

$$\Phi_t = \sum_j \int_{-\infty}^{+\infty} (\mathbb{E}_t V_{jt+1} + \nu \epsilon_{jt}) f(\epsilon_{jt}) \prod_{l \neq j} F(\mathbb{E}_t V_{jt+1} - \mathbb{E}_t V_{lt+1} + \epsilon_{jt}) d\epsilon_{jt}.$$

Substituting for $F(\epsilon)$ and $f(\epsilon)$, we have

$$\Phi_t = \sum_j \int_{-\infty}^{+\infty} (\mathbb{E}_t V_{jt+1} + \nu \epsilon_{jt}) e^{(-\epsilon_{jt} - \overline{\gamma})} e^{\left[-\exp(-\epsilon_{jt} - \overline{\gamma}) \sum_l \exp(-\frac{1}{\nu}(\mathbb{E}_t V_{jt+1} - \mathbb{E}_t V_{lt+1})) \right]} d\epsilon_{jt}$$

Using $\lambda_t \equiv \log \sum_j \exp\{-\frac{1}{v}(\mathbb{E}_t V_{jt+1} - \mathbb{E}_t V_{lt+1})\}$ and the change of variable $\xi_{jt} = \epsilon_{jt} + \overline{\gamma} - \lambda_t$, this yields

$$\Phi_t = \sum_j \exp(-\lambda_t)[\mathbb{E}_t V_{jt+1} + \nu \epsilon_{jt}]$$

Substitute the representation of λ_t, we have

$$\Phi_t = \nu \left(\log \sum_j \exp\left(\frac{1}{\nu} \mathbb{E}_t V_{jt+1} \right) \right)$$

and therefore, the current value in region j can be represented by

$$V_{jt} = u(c^{\star}_{jt}, n^{\star}_{jt}, h^{\star}_{jt}) + \kappa_j + \beta\psi\nu \left(\log \sum_l \exp\left(\frac{1}{\nu} \mathbb{E}_t V_{lt+1} \right) \right) + \beta(1-\psi)\mathbb{E}_t V_{jt+1}.$$

Q.E.D.

A.4 Regional and Aggregate Dynamic Systems

This subsection shows the aggregation result and the regional and aggregate dynamic systems that represent the behavior of regional and aggregate variables in a locally-perturbed economy around steady state.

A.4.1 The Regional Dynamic System

In the empirical investigation, the regional impulse responses are estimated by an instrumental variable local projection method for panel data. The baseline regressions include region and time fixed effects soaking up all region-specific time-invariant terms as well as aggregate time series variation. Therefore, the regional impulse responses are inferred from differential responses of local variables between two regions.

In order to make the model-base regional impulse responses conceptually consistent with the estimates, I log-linearize the equilibrium system around steady state[31] and difference the whole system between two regions, where the differential regional variables are denoted by a lower-case letter with tildes. Therefore, the aggregate variables are differenced out as in the fixed-effect regressions for panel data. This yields a regional dynamic system only including regional variables $\{\tilde{\hat{y}}_{jt}, \tilde{\hat{c}}_{jt}, \tilde{\hat{i}}_{jt}, \tilde{\hat{x}}_{jt}, \tilde{\hat{g}}_{jt}, \tilde{\hat{m}}_{jt}, \tilde{\hat{n}}_{jt}, \tilde{\hat{P}}_{jt}, \tilde{\hat{W}}_{jt}, \tilde{\hat{V}}_{jt}, \tilde{\hat{\varphi}}_{jt}, \tilde{\hat{\mu}}_{jt}, \tilde{z}_{jt}\}$, shown as

$$y^\star \tilde{\hat{y}}_{jt} = c^\star \tilde{\hat{c}}_{jt} + i^\star \tilde{\hat{i}}_{jt} + x^\star \tilde{\hat{x}}_{jt} + g^\star \tilde{\hat{g}}_{jt}, \tag{2.12}$$

$$\tilde{\hat{y}} = \alpha \tilde{\hat{m}}_{jt} + (1-\alpha)\tilde{\hat{n}}_{jt}, \tag{2.13}$$

$$\tilde{\hat{i}}_{jt} = -\eta \tilde{\hat{P}}_{jt}, \tag{2.14}$$

$$c^\star \tilde{\hat{c}}_{jt} = w^\star n^\star (\tilde{\hat{W}}_{jt} - \tilde{\hat{P}}_{jt} + \tilde{\hat{n}}_{jt}), \tag{2.15}$$

$$\phi_n \tilde{\hat{n}}_{jt} = (\tilde{\hat{W}}_{jt} - \tilde{\hat{P}}_{jt}) - \sigma \tilde{\hat{c}}_{jt}, \tag{2.16}$$

$$\tilde{\hat{P}}_{jt} = (1-\alpha)(\tilde{\hat{m}}_{jt} - \tilde{\hat{n}}_{jt}), \tag{2.17}$$

$$\tilde{\hat{W}}_{jt} - \tilde{\hat{P}}_{jt} = \alpha(\tilde{\hat{m}}_{jt} - \tilde{\hat{n}}_{jt}), \tag{2.18}$$

$$V^\star \tilde{\hat{V}}_{jt} = (c^\star)^{1-\sigma}\tilde{\hat{c}}_{jt} - \lambda_n (n^\star)^{1+\phi_n}\tilde{\hat{n}}_{jt} + \lambda_h (h^\star)^{1-\sigma_h}\tilde{\hat{h}}_{jt} + \beta(1-\psi)V^\star \mathbb{E}_t \tilde{\hat{V}}_{jt+1}, \tag{2.19}$$

$$\tilde{\hat{\varphi}}_{jt} = \frac{V^\star}{\nu}\mathbb{E}_t \tilde{\hat{V}}_{jt+1}, \tag{2.20}$$

$$\tilde{\hat{\mu}}_{jt+1} = (1-\psi)\tilde{\hat{\mu}}_{jt} + \psi \tilde{\hat{\varphi}}_{jt}, \tag{2.21}$$

$$\tilde{\hat{h}}_{jt+1} = (1-\delta_h)\tilde{\hat{h}}_{jt} + \delta_h \zeta \tilde{\hat{x}}_{jt} - \tilde{\hat{\mu}}_{jt+1} + \tilde{\hat{\mu}}_{jt}, \tag{2.22}$$

$$\tilde{\hat{g}}_{jt} = \tilde{z}_{jt}, \tag{2.23}$$

$$\tilde{z}_{jt} = \rho_z \tilde{z}_{jt-1} + \tilde{\varepsilon}^z_{jt}, \tag{2.24}$$

$$\tilde{\hat{x}}_{jt+1} = (1-\phi_h)\tilde{\hat{x}}_{jt} + (\phi_h - \psi)\zeta^{-1}(\tilde{\hat{\varphi}}_{jt} - \tilde{\hat{\mu}}_{jt}), \tag{2.25}$$

[31] Notably, the labor distribution $\{\mu_j^\star\}_{j\in\mathcal{J}}$ and individual's value in regions $\{V_j^\star\}$ at steady state depend on local amenity $\{\kappa_j\}$.

where (2.12) is the regional final goods clearing condition, (2.13) is the final goods production function, (2.14) is the investment goods demand, (2.15) is the household's budget constraint, (2.16) the labor-consumption optimality condition, (2.17) and (2.18) are firm's input decisions, (2.19) is the Bellman equation, (2.20) is the optimal migration policy, (2.21) is the evolution of population distribution, (2.22) is the evolution of regional housing per resident, (2.25) is the evolution of construction spending subject to completion lags.

In equilibrium, individual agents guess the law of motion on regional population. Given this law of motion, individual agents solve their own optimal migration policies. Rational expectation implies that the realized evolution of regional population under individual's optimal migration policy should be consistent with the perceived law of motion.

Denote regional state variables by $\hat{\mathtt{x}}_{jt} = (\hat{x}_{jt}, z_{jt}, \hat{h}_{jt}, \hat{\mu}_{jt})'$. The solution for migration policy is given by $\hat{\varphi}_{jt} = \Theta' \hat{\mathtt{x}}_{jt}$ and the regional state variables can be represented as follows, $\hat{\mathtt{x}}_{jt} = \Phi_1 \hat{\mathtt{x}}_{jt-1} + \epsilon_{jt}$, where $\Theta = (\theta_1, .., \theta_4)'$ and Φ_1 are the coefficients solved from the above regional dynamic system.

The expected future value can be represented by a linear approximation around symmetric steady state with respect to aggregate and region-specific state variables, $\hat{S}_t$ and $\hat{\mathtt{x}}_{jt}$, $\mathbb{E}_t V_j(\mathbf{s}'|\mathbf{s}) \approx V_j^{ss} + \theta_s \hat{S}_t + \theta_x \hat{\mathtt{x}}_{jt}$, where θ_s and θ_x are the first-order derivatives with respect to aggregate state variables $\hat{S}_t$ and regional state variables $\hat{\mathtt{x}}_{jt}$ valued at symmetric steady state. Under Gumbel distribution assumption, the optimal relocation policy can be shown as

$$\varphi_{jt}^{\star} = \frac{\exp[\mathbb{E}_t V_j(\boldsymbol{s}_{t+1})]^{\frac{1}{\nu}}}{\sum_l \exp[\mathbb{E}_t V_l(\boldsymbol{s}_{t+1})]^{\frac{1}{\nu}}} \approx \frac{\exp(V_j^{\star})^{\frac{1}{\nu}} \exp(\theta_s \hat{S}_t)^{\frac{1}{\nu}} \exp(\theta_z \hat{\mathtt{x}}_{jt})^{\frac{1}{\nu}}}{\sum_l \exp(V_l^{\star})^{\frac{1}{\nu}} \exp(\theta_s \hat{S}_t)^{\frac{1}{\nu}} \exp(\theta_z \hat{\mathtt{x}}_{lt})^{\frac{1}{\nu}}} = \frac{\mu_j^{\star} \exp(\Theta' \hat{\mathtt{x}}_{jt})}{\sum_l \mu_l^{\star} \exp(\Theta' \hat{\mathtt{x}}_{lt})},$$

where Θ is the coefficients solved from the regional dynamic system.

A.4.2 The Aggregate Dynamic System

Flexible Perturbation. Given the discussion in Section 3, regional reallocation gives rise to aggregate impacts through a covariance effect. This nonlinear aggregate effect washes out in the standard first-order log-linearization approximation. To characterize the covariance effects in a tractable way, I apply a flexible perturbation method as follows. For any aggregate variable that is defined by the sum of per-resident regional counterparts weighted by regional populations, e.g., aggregate consumption $C_t = \sum_j c_{jt}\mu_{jt}$, and due to the fact that population shares across regions add up to one, $\sum_j \mu_{jt} = 1$, we have the following identity,

$$\frac{C_t - c^\star}{c^\star} = \frac{\sum_j c_{jt}\mu_{jt} - \sum_j c^\star \mu_{jt}}{c^\star} = \Sigma_j(\frac{c_{jt} - c^\star}{c^\star})\mu_{jt}. \tag{2.26}$$

Therefore, $\hat{\mathcal{C}}_t = \sum_j \hat{c}_{jt}\mu_{jt}$, implying that the cross-sectional inner-products between local variable exposures and population shares are equivalent to the aggregate counterparts in log-deviation from steady state. In addition, expanding μ_{jt} around steady state, i.e., $\mu_{jt} = (1+\hat{\mu}_{jt})\mu_j^\star$, the log-deviation of aggregate consumption can be shown by

$$\hat{\mathcal{C}}_t = \sum_j \hat{c}_{jt}\mu_j^\star + \sum_j \hat{c}_{jt}\hat{\mu}_{jt}\mu_j^\star. \tag{2.27}$$

This equation indicates that the second-order term $\sum_j \hat{c}_{jt}\hat{\mu}_{jt}\mu_j^\star$ captures the covariance effect arising from regional reallocation. The major advantage of such a flexible perturbation method is that it enables us to characterize the nonlinear covariance effect in a linear manner, which significantly reduces computational intensity but preserves the covariance effect.

Aggregate Dynamic System. Aggregating up the locally-perturbed equilibrium conditions weighted by regional populations, the aggregate dynamics in the benchmark model can be represented by the following aggregate dynamic system including aggregate variables $\{\hat{\mathcal{Y}}_t, \hat{\mathcal{C}}_t, \hat{\mathcal{I}}_t, \hat{\mathcal{G}}_t, \hat{\mathcal{X}}_t, \hat{\mathcal{K}}_t, \hat{\mathcal{N}}_t, \hat{R}_t, \hat{P}_t, \hat{\mathcal{W}}_t, \hat{R}^k_{t+1}, \hat{Q}_t, \hat{Q}^\star_t, \hat{\mathcal{Z}}_t\}$,

$$y^\star \hat{\mathcal{Y}}_t = c^\star \hat{\mathcal{C}}_t + i^\star \hat{\mathcal{I}}_t + g^\star \hat{\mathcal{G}}_t + x^\star \hat{\mathcal{X}}_t, \tag{2.28}$$

$$\hat{\mathcal{Y}}_t = \alpha \hat{\mathcal{K}}_t + (1-\alpha)\hat{\mathcal{N}}_t, \tag{2.29}$$

$$\hat{K}_{t+1} = (1-\delta)\hat{K}_t + \delta \hat{\mathcal{I}}_t, \tag{2.30}$$

$$-\sigma \hat{\mathcal{C}}_t = \mathbb{E}_t \hat{R}_{t+1} - \sigma \mathbb{E}_t \hat{\mathcal{C}}_{t+1} + \hat{\hat{P}}_t - \mathbb{E}_t \hat{\hat{P}}_{t+1}, \tag{2.31}$$

$$\phi_n \hat{\mathcal{N}}_t = -\sigma \hat{\mathcal{C}}_t + (\hat{\mathcal{W}}_t - \hat{\mathcal{P}}_t) + \hat{\mathcal{N}}_t, \tag{2.32}$$

$$\mathbb{E}_t \hat{R}_{t+1} = \frac{R^{k\star}}{R^{k\star} + 1 - \delta} \mathbb{E}_t \left(\hat{R}^k_{t+1} - \hat{P}_{t+1} \right) + \mathbb{E}_t \hat{\pi}_{t+1}, \tag{2.33}$$

$$\hat{Q}_t = \hat{\mathcal{P}}_t - (1-\alpha)\hat{\mathcal{K}}_t + (1-\alpha)\hat{\mathcal{N}}_t, \tag{2.34}$$

$$\hat{\mathcal{W}}_t = \hat{\mathcal{P}}_t + \alpha \hat{\mathcal{K}}_t - \alpha \hat{\mathcal{N}}_t, \tag{2.35}$$

$$\hat{Q}^\star_t = (1-\beta\theta)\hat{R}^k_t + \beta\theta \mathbb{E}_t \hat{Q}^\star_{t+1}, \tag{2.36}$$

$$\hat{Q}_t = \theta \hat{Q}_{t-1} + (1-\theta)\hat{Q}^\star_t, \tag{2.37}$$

$$\hat{R}_t = \phi_\pi (\hat{P}_t - \hat{P}_{t-1}), \tag{2.38}$$

$$\hat{\mathcal{G}}_t = \hat{\mathcal{Z}}_t, \tag{2.39}$$

$$\hat{\mathcal{Z}}_t = \sum_j z_{jt} \mu_{jt}, \tag{2.40}$$

$$\hat{\mathcal{X}}_t = \sum_j \hat{x}_{jt} \mu_{jt}, \tag{2.41}$$

with localized government spending processes, $\hat{z}_{jt} = \rho_z \hat{z}_{jt-1} + \epsilon^z_{jt}$, law of motion in local construction spending, $\hat{x}_{jt+1} = (1-\phi_h)\hat{x}_{jt} + (\phi_h - \psi)\zeta^{-1}(\hat{\varphi}_{jt} - \hat{\mu}_{jt})$, evolution of population distribution, $\hat{\mu}_{jt+1} = (1-\psi)\hat{\mu}_{jt} + \psi \hat{\varphi}_{jt}$, evolution of local housing stock per resident, $\hat{h}_{jt+1} = (1-\delta_h)\hat{h}_{jt} + \delta_h \zeta \hat{x}_{jt} - \hat{\mu}_{jt+1} + \hat{\mu}_{jt}$, and optimal migration policy,

$\hat{\varphi}_{jt} = \Theta' \hat{\mathtt{x}}_{jt}$, where $\Theta = (\theta_1, .., \theta_4)'$ can be solved from the regional dynamic system and $\hat{\mathtt{x}}_{jt} = (\hat{x}_{jt}, z_{jt}, \hat{h}_{jt}, \hat{\mu}_{jt})'$ denotes regional economic state variables.

A.4.3 Proof of Proposition 3 in Section 2.4

Proposition 3 (*Aggregate Dynamics*) *The behavior of aggregate variables in the locally-perturbed economy is identical to that of an augmented representative economy in the form of a linear dynamic system* $\Gamma_0 \mathbb{E}_t \hat{\mathcal{Y}}_{t+1} = \Gamma_1 \hat{\mathcal{Y}}_t + \Gamma_2 \tilde{\mathcal{X}}_t$, *where* Γ_0, Γ_1, *and* Γ_2 *are matrices of coefficients,* $\hat{\mathcal{Y}}_t = \sum_j \hat{\mathtt{y}}_{jt} \mu_{jt}$ *stands for aggregate variables and* $\tilde{\mathcal{X}}_t = \sum_j \hat{\mathtt{x}}_{jt} \tilde{\varphi}_{jt}$ *stands for reallocation summary statistics.*

Proof: Aggregating up the locally-perturbed equilibrium conditions by regional populations, the aggregate dynamics in the benchmark model can be represented by a linear dynamic system consisting of equations (2.28) to (2.41). By solving the regional dynamic system, the regional economic state variables can be represented as follows, $\hat{\mathtt{x}}_{jt} = \Phi_1 \hat{\mathtt{x}}_{jt-1} + \epsilon_{jt}$, and the optimal migration policy is $\varphi_{jt} = \frac{\mu_j^\star \exp(\Theta' \hat{\mathtt{x}}_{jt})}{\sum_l \mu_l^\star \exp(\Theta' \hat{\mathtt{x}}_{lt})}$ and the evolution of regional population is $\mu_{jt+1} = (1-\psi)\mu_{jt} + \psi \varphi_{jt}$, where Φ_1 and Θ can be solved from the regional dynamic system of (2.12) to (2.25) and $\hat{\mathtt{x}}_{jt} = (\hat{x}_{jt}, z_{jt}, \hat{h}_{jt}, \hat{\mu}_{jt})'$ denotes regional state variables. Denote $\hat{\boldsymbol{X}}_t = \sum_j \hat{\mathtt{x}}_{jt} \mu_{jt}$ as "aggregate counterparts of regional state variables." Plugging in the evolution of regional population and regional state variables, this yields

$$\begin{aligned}
\hat{\boldsymbol{X}}_t = \sum_j \hat{\mathtt{x}}_{jt} \mu_{jt} &= \sum_j (\Phi_1 \hat{\mathtt{x}}_{jt-1} + \epsilon_{jt})(\mu_{jt-1} + \psi \tilde{\varphi}_{jt-1}) \\
&= \Phi_1 \sum_j \hat{\mathtt{x}}_{jt-1} \mu_{jt-1} + \psi \Phi_1 \sum_j \hat{\mathtt{x}}_{jt-1} \tilde{\varphi}_{jt-1} + \sum_j \epsilon_{jt} \mu_{jt}.
\end{aligned}$$

Thus, the evolution of aggregate counterparts of regional state variables,

$$\hat{\boldsymbol{X}}_t = \Phi_1 \hat{\boldsymbol{X}}_{t-1} + \Phi_2 \tilde{\mathcal{X}}_{t-1} + \mathcal{E}_t,$$

where $\tilde{\mathcal{X}}_t = \sum_j \hat{\mathrm{x}}_{jt} \tilde{\varphi}_{jt}$ is a function of regional state variables $\hat{\mathrm{x}}_{jt} = (\hat{x}_{jt}, z_{jt}, \hat{h}_{jt}, \hat{\mu}_{jt})'$.

Therefore, the behavior of aggregate variables in the locally-perturbed economy can be represented by a linear dynamic system $\Gamma_0 \mathbb{E}_t \hat{\mathcal{Y}}_{t+1} = \Gamma_1 \hat{\mathcal{Y}}_t + \Gamma_2 \tilde{\mathcal{X}}_t$, where Γ_0, Γ_1, and Γ_2 are matrices of coefficients, $\hat{\mathcal{Y}}_t = \sum_j \hat{\mathrm{y}}_{jt} \mu_{jt}$ stands for aggregate variables and $\tilde{\mathcal{X}}_t = \sum_j \hat{\mathrm{x}}_{jt} \tilde{\varphi}_{jt}$ stands for reallocation summary statistics.

Q.E.D.

A.4.4 Proof of Proposition 4 in Section 2.4

Proposition 4 (*Regional Dynamics*) *The behavior of regional variables in the log-linearized economy is identical to that of a regional dynamic economy in the form of* $\Phi_0 \mathbb{E}_t \tilde{\mathbf{y}}_{t+1} = \Phi_1 \tilde{\mathbf{y}}_t + \Phi_2 \tilde{\mathbf{z}}_t$, *where* Φ_0, Φ_1, *and* Φ_2 *are matrices of coefficients,* $\tilde{\mathbf{y}}_t$ *stands for the differential regional variables,* $\tilde{\mathbf{z}}_t$ *represents regional idiosyncratic shocks.*

Proof: The regional dynamic system of (2.12) to (2.25) is a linear dynamic system that represents the behavior of differential regional variables in the log-linearized economy.

Q.E.D.

B Details on Quantitative Analysis

B.1 The Solution Method

This subsection describes the numerical algorithm to solve both the regional and aggregate dynamics of the model. The algorithm includes three steps.

- Step 1 solves the regional dynamic system of (2.12) to (2.25). This yields the solution for optimal migration policy as well as the evolution of regional population and other regional state variables.
- Step 2 represents the evolution of "reallocation summary statistics" to be a linear difference system, given a vector of regional shocks.
- Step 3 solves the aggregate dynamic system of (2.28) to (2.41) by plugging in the evolution of reallocation summary statistics.

Step 1. Solving the Regional Dynamic System

The regional dynamic system is given by equations (2.12) to (2.25), where the regional state variables are denoted by $\hat{\mathtt{x}}_{jt} = (\hat{x}_{jt}, z_{jt}, \hat{h}_{jt}, \hat{\mu}_{jt})'$. While individual agents make their own optimal migration policies, their decisions also depend on the expectations about others migration decisions. In equilibrium, individual agents guess the law of motion of regional population and the realized evolution of regional population should be consistent with their perceived law of motion.

Guess a solution for optimal migration policy and substitute it into the evolution of regional construction spending. Solve the regional dynamic system and obtain the solution to optimal migration policy. Update the guess of the solution for optimal migration policy such that the guessed solution is consistent with the realized solution.

Step 2. Representing the Reallocation Summary Statistics

Denote "aggregate counterparts of regional state variables" by $\hat{\boldsymbol{X}}_t = \sum_j \hat{\mathtt{x}}_{jt}\mu_{jt}$. Plugging in the evolution of regional population and regional state variables, this yields

$$\begin{aligned}\hat{\boldsymbol{X}}_t = \sum_j \hat{\mathtt{x}}_{jt}\mu_{jt} &= \sum_j (\Phi_1 \hat{\mathtt{x}}_{jt-1} + \epsilon_{jt})[\mu_{jt-1} + \psi\tilde{\varphi}_{jt-1}] \\ &= \Phi_1 \sum_j \hat{\mathtt{x}}_{jt-1}\mu_{jt-1} + \psi\Phi_1 \sum_j \hat{\mathtt{x}}_{jt-1}\tilde{\varphi}_{jt-1} + \sum_j \epsilon_{jt}\mu_{jt}.\end{aligned}$$

Thus, the evolution of aggregate counterparts of regional state variables,

$$\hat{\boldsymbol{X}}_t = \Phi_1 \hat{\boldsymbol{X}}_{t-1} + \psi\Phi_1 \tilde{\mathcal{X}}_{t-1} + \mathcal{E}_t,$$

where $\tilde{\mathcal{X}}_t = \sum_j \hat{\mathtt{x}}_{jt}\tilde{\varphi}_{jt}$ is a function of regional state variables $\hat{\mathtt{x}}_{jt} = (\hat{x}_{jt}, z_{jt}, \hat{h}_{jt}, \hat{\mu}_{jt})'$ and $\varphi_{jt} = \frac{\mu_j^\star \exp(\Theta'\hat{\mathtt{x}}_{jt})}{\sum_l \mu_l^\star \exp(\Theta'\hat{\mathtt{x}}_{lt})}$. Given a vector of regional shocks $\{\varepsilon_{jt}^z\}_{j\in\mathcal{J}}$, the solution for optimal migration policy and the evolution of regional state variables yield a set of simulated data of the aggregate counterparts of regional state variables $\{\hat{\boldsymbol{X}}_t\}$. According to the discussion about covariance effects in Section 3, the covariance effects wash out in the first-order approximation. In order to capture the covariance effects, I use the mean and the dispersion of regional government spending by $z_t^m = \frac{1}{J}\sum_j z_{jt}$ and $z_t^d = \frac{1}{J}\sum_j z_{jt}^2$ under the shocks $\{\varepsilon_{jt}^z\}_{j\in\mathcal{J}}$ to approximate the evolution of reallocation summary statistics $\{\tilde{\mathcal{X}}_t\}$. Therefore, the process of $\hat{\boldsymbol{X}}_t$ can be approximated by $\hat{\boldsymbol{X}}_t = \Phi_1\hat{\boldsymbol{X}}_{t-1} + A_1 z_t^m + A_2 z_t^d$, where $z_t^m = \rho_z z_{t-1}^m + \epsilon_t^m$ and $z_t^d = \rho_z^2 z_t^d + \epsilon_t^d$. The coefficients A_1 and A_2 can be estimated by running a regression of simulated data$\{\hat{\boldsymbol{X}}_t - \Phi_1\hat{\boldsymbol{X}}_{t-1}\}$ over $\{z_t^m, z_t^d\}$. Thus, the evolution of aggregate counterparts of regional state variables can be represented by a linear difference system.

Step 3. Solving the Aggregate Dynamic System

Knowing the representation of processes of $\hat{\mathcal{Z}}_t$ and $\hat{\mathcal{X}}_t$ given a vector of regional shocks $\{\varepsilon^z_{jt}\}_{j\in\mathcal{J}}$ as a linear system of $\{\hat{\boldsymbol{X}}_{t-1}, z^m_t, z^d_t\}$, we can solve the aggregate dynamic system of (2.28) to (2.41) by using standard methods to solve a linear rational expectation system. This yields the aggregate impulse response functions to a vector of regional shocks.

B.2 Quantitative Robustness

B.2.1 Model Fitness

Regional Construction and Immigration Shocks. Imposing an immigration shock of one percent in regional population, the impulse responses of regional housing construction are shown in Figure 2.14. The left panel displays the state-level construction spending in my calibrated model. The right panel displays the impulse responses of MSA-level construction employment rate estimated by Howard (2018).

Estimated vs. Model-based Regional Impulse Responses of Localized Government Spending. Figure 2.15 plots the impulse responses to regional real military spending per resident over various horizons. The blue circle denotes the point estimates while the black dot represents the model-based impulse responses of state-level government spending with AR(1) coefficient of 0.75 at an annual frequency, which is equivalent to Nakamura and Steinsson's (2014) estimate of 0.933 at a quarter frequency.

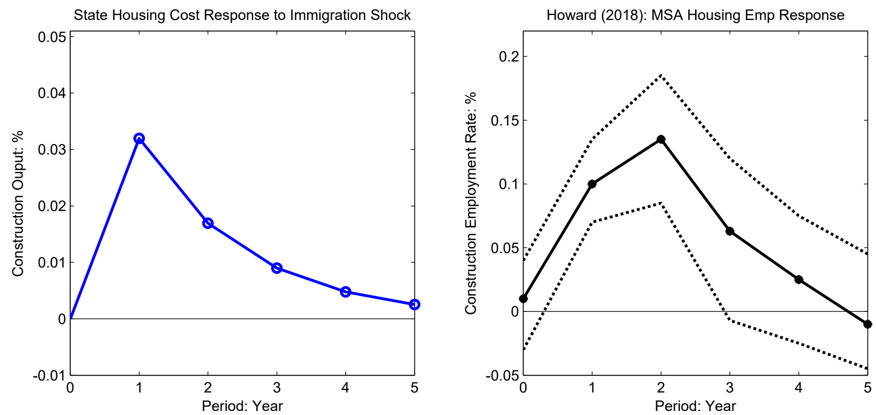


Figure 2.14: Regional Construction Responses to Immigration Shocks

Notes: The left panel displays the impulse responses of state-level construction spending to an immigration shock of 1 percent regional population in my calibrated model. The right panel is the impulse responses of MSA-level construction employment rate estimated by Howard (2018) (Figure 8 in Page 22) with 95 percent confidence interval where standard errors are clustered by state.

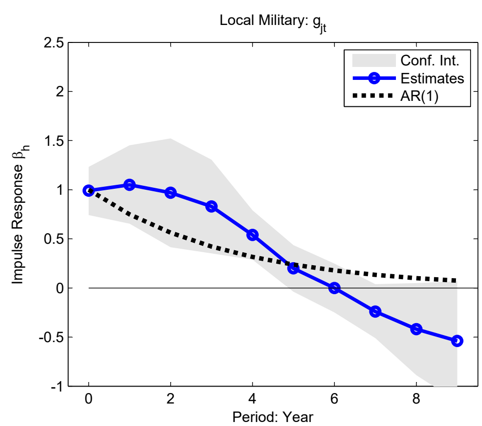


Figure 2.15: Estimated Impulse Response of Regional Military Spending

Notes: The circles are LP-IV estimates of regional impulse responses of regional military spending to a national military buildup. The gray area is 95% confidence interval. Standard errors are robust and clustered by state. The sample period is 1966-2006. The dotted line is the AR(1) process with persistence coefficient 0.75 at an annual frequency.

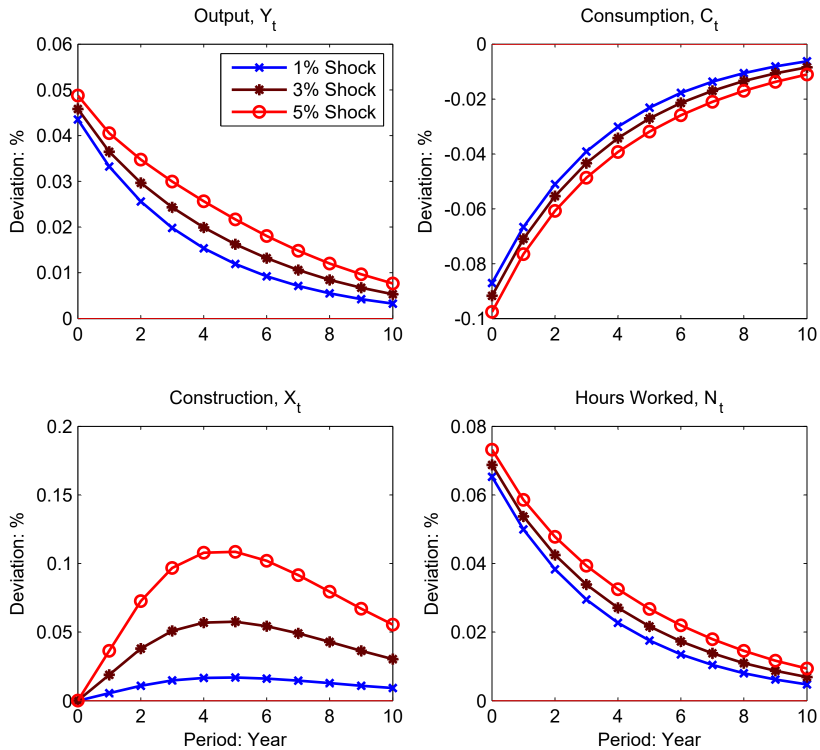


Figure 2.16: Aggregate Impacts of Locally-targeted Fiscal Policies

Notes: The figure reports the impulse responses normalized by the initial size of the shocks. In particular, the size-normalized impulse response of an $x\%$ shock is the corresponding impulse responses divided by x.

B.2.2 The Impacts of Locally-Targeted Fiscal Policies

This subsection analyzes the impacts of locally-targeted fiscal policies in which the government purchases regional products in one particular region. For instance, the federal government builds a new administration center in Maryland or launches a highway upgrading program in Texas. The existing literature based on linearized models without regional reallocation predicts that the aggregate impacts of a localized government spending shock are proportional (or linear) to its size. I define a *size-normalized* impulse response of an x percent shock by the impulse responses divided by x. Therefore, existing linearized models predict the same normalized impulse responses regardless of the size of the shock.

Taking regional reallocation into account, I apply the calibrated model to quantify the impacts of a localized government spending shock in 1 percent, 3 percent, and 5 percent in California where the steady state population is 10 percent. Figure 2.16 displays the size-normalized impulse responses with respect to 1 percent, 3 percent, and 5 percent shocks. As the size of the shock gets larger, aggregate output, consumption, housing construction, and hours worked respond at a greater rate than the size of localized government spending expands. Take the 1 percent and 3 percent shock as an example. The aggregate output in response to a 3 percent shock increases more than three times of the aggregate output response to a 1 percent shock. Intuitively, if labor reallocation is allowed across regions, larger regional shocks lead to a larger scale of net directed migration. Given a fixed amount of land, labor migrates to booming regions and the regional housing construction increases more than the increase of population. Therefore, the aggregate construction rises in response to a larger scale of labor reallocation (Panel c) and the increase of aggregate construction serves as a demand shifter and reduces aggregate consumption even more (Panel b). As a result, both aggregate hours worked (Panel d) and aggregate output (Panel a) increase even more.

Chapter 3

Microeconomic Heterogeneity and Macroeconomic Aggregation

> "To me, the most extraordinary thing regarded historically, is the complete disappearance of the theory of the demand and supply for output *as a whole*, [...] after it had been for a quarter of a century the most discussed thing in economics."
>
> John Maynard Keynes (1936)

3.1 Introduction

The behavior of aggregate economy *as a whole* stems from a large amount of heterogeneous microeconomic units and their interplay in the market. In order to understand how an aggregate economy behaves, economists should, in principle, work out the theories not only of how microeconomic units behave but also of how to aggregate these microeconomic responses appropriately. An extensive body of literature studies how to add micro production or demand functions into their aggregate counterparts. Such methods exist, but they rely on assumptions that one should hardly expect to be approximated in actual economies.[1]

[1] For example, Leontief (1947), Nataf (1948), Gorman (1953), and Fisher (1969) provide necessary and sufficient conditions for aggregation. However, these conditions are in fact stringent; for a comprehensive review, see Felipe and Fisher (2003).

Instead, modern macroeconomics turns to a different sort of approach—the representative agent model—that works out theories how an individual behaves and transfers these rules of behavior to the aggregate level. The neoclassical production function and the aggregate consumption function are famous examples. This approach enables us to deliver some useful aggregate relationships that can be empirically identified or tested without knowing disaggregate details, while this approach also shuts down the possibility of studying how micro heterogeneity matters for macro impacts. It turns out that a recent literature builds on disaggregate models and sheds new light on the aggregate implications of micro heterogeneity (e.g., Auclert, 2019; Baqaee and Farhi, 2019). These studies apply aggregation to derive useful sufficient statistics that can be identified or estimated with some but not all disaggregate details. However, most of these studies focus on the macro impact of transitory shocks or comparative statics analysis. It is still an open question to what extent aggregation can be implemented to analyze dynamic macro impacts in a disaggregate, recursive stochastic dynamic context.

In this paper, we tackle the aggregation issue in the context of recursive stochastic dynamic models and propose a new method to study the macro impact of micro heterogeneity. We show that, after differentiating disaggregate equilibrium conditions, aggregation can be admitted to construct a recursive stand-in macro system that displays the same aggregate dynamics as in the disaggregate model. This method hinges on an insight that linearity makes aggregation tractable and convenient. The reason why exact aggregation is difficult is because microeconomic responses are in general nonlinearly interlaced in equilibrium conditions. However, as these nonlinear equilibrium conditions can be decomposed by approximation order in a linear manner, aggregation can be implemented as a linear combination and thus becomes convenient. We bring this insight to a broad class of dynamic stochastic general equilibrium (DSGE) models with heterogeneous micro structures. This method provides a sufficient-statistic characterization of macro shocks and aggregate relationships in terms of heterogeneous micro shocks and structures. The first- and second-order macro shocks can be sharply characterized by the average and the dispersion of micro counterparts weighted by their impact intensities. We apply this method in a variety of applications to illustrate the importance of micro heterogeneity and nonlinearity in macroeconomics as well as how to implement it in different contexts.

In Section 3.2, we start with an illustrative example of production allowing for rich sectoral or locational heterogeneity to i) show how to find the stand-in macro model linking a list of macro variables to micro shocks but not the micro variables and ii) highlight economic lessons about the macro impact of micro heterogeneity. The micro production uses capital and labor as inputs with constant return to scale property and allows for arbitrary functional forms, e.g., micro production functions may be not additively separable between capital and labor. As pointed out by the early literature on aggregation, it is unlikely to have a consistent aggregate production function that exactly captures the aggregate behavior of output and factor inputs. However, it is still plausible to construct a first-order approximate stand-in model, since micro production functions turn to be log-linear in the first-order approximate model and thus Gorman's aggregation conditions can be satisfied. The first lesson is that, to the first order, the implications of the micro model with sectoral or locational heterogeneity are the same as those of a representative framework, only with economy-wide factor shares and aggregate productivity playing the key role but without knowing all disaggregate details. Although the aggregate production function generally holds to the first-order, it is not legitimate to directly assume a representative firm with optimizing behavior under the aggregate production function. This is because of the second lesson that, even if there exists a unique inverse relationship between factor intensity and relative return rates of factor for all micro production function to the first order, there may not be a unique inverse relationship at the aggregate level in the presence of micro production heterogeneity. Put it another way, the aggregate factor demand derived from a representative firm with an aggregate production function is barely equivalent to the aggregate factor demand derived from a disaggregate model. We show that such deviation between micro and macro levels can be sufficiently captured by a new but unfamiliar shock, which comes from the interaction of fixed heterogeneity in the micro production structures and heterogeneity in the micro shocks. Allowing for rich heterogeneity in micro production structures, a research can make use of micro equilibrium conditions to derive a stand-in macro model that links the behavior of macro variables to a set of aggregate shocks as sufficient statistics, which potentially are averages of micro shocks weighted by their corresponding impact intensities.

Although it is relatively easy to find the first-order approximate stand-in model owing to the linearity of first-order approximation, aggregation can be difficult in the second- and higher-order approximation. This is because, under standard Taylor expansion, the variables in higher-order approximate model are quadratic functions instead of linear functions. To overcome this difficulty, we propose a technique to decompose the second- and higher-order terms via differentials in a linear manner. For instance, the second-order deviation of a variable can be decomposed into its first- and second-order differentials. In the second-order system, the quadratic terms are only associated with the cross-product of first-order differentials but have nothing to do with the second-order differentials. Under this decomposition, the second-order differentials are linear in the second-order approximate system. This method is also related to the literature of solving rational expectations models by using higher-order approximation via either pruning terms of higher-order components than the considered approximation order or via differentials (e.g., Lombardo and Sutherland, 2007; Kim, Kim, Schaumburg, and Sims, 2008; Johnston, King, and Lie, 2014). Our method applies familiar linear rational expectations techniques to successively solve for higher-order approximation. Differently, we utilize this approach in the context allowing for aggregating micro heterogeneity. The approximation by order yields a linear representation and thus provides tractability for aggregation over micro equilibrium conditions by approximation order. Using this method, we show how to find the second-order stand-in model that links a list of macro variables to the first- and second-order macro shocks, which can be shaped by the average and the dispersion of the micro counterparts weighted by their micro impact intensities.

In Section 3.3, we apply this method in a broad class of DSGE models with micro heterogeneity. In principle, adding up microeconomic behavior into the macro level is very restrictive, because microeconomic behavior can be nonlinearly interlaced and distribution evolution can be interdependent across microeconomic units in equilibrium. Two methods are prevailing in the modern macroeconomics literature: the representative-agent approach and the numerical aggregation approach (e.g., applications of Krusell and Smith (1997) algorithm in disaggregate models with household's or firm's heterogeneity). Although the relationships among aggregates can be studies by the representative-agent approach, this approach is almost silent about how micro heterogeneity affects aggregate responses. The numerical aggregation approach solves macro models with heterogeneous micro units by using numerical methods. This approach can generate quantitative aggregate impacts of micro heterogeneity, but those quantitative results are usually not easy to interpret qualitatively in the sense what types of micro heterogeneity matter and through what channels. Our method combine strengths of existing approaches. Using our method, aggregation becomes convenient because microeconomic equilibrium conditions are decomposed by approximation order in a linear manner. To each order, the stand-in macro system is functionally equivalent to a representative-agent model in which macro elasticities and shocks are explicitly characterized in terms of micro elasticities and shocks. These macro elasticities and shocks can be considered as sufficient statistics that characterize the macro impact of micro heterogeneity.

In Section 3.4 and 3.5, we apply this method in a variety of applications to illustrate the importance of micro heterogeneity and nonlinearity in macroeconomics as well as how to implement it in different contexts. Most of existing studies on the macro impact of micro heterogeneity using aggregation focus on the macro impact of transitory shocks or comparative statistic analysis. In particular, Baqaee and Farhi (2019) provide a sufficient-statistic characterization on the second-order macro impact of micro productivity via input-output network. Although their characterization is a comparative statics analysis, it can be extended to a dynamic context by re-indexing goods over time as in Arrow-Debreu framework. Thus, these sufficient statistics become time-variant and conditional on the sequence of realizations of all past and present random variables and shocks. This analysis is difficult to be applied by standard time series methods, because their characterizations lack of a recursive structure. One alternative approach proving useful is the recursive equilibrium analysis (Prescott and Mehra, 1980). Applying our method, we can provide an alternative characterization on the macro impact of micro shocks by a recursive stochastic dynamic aggregate system. The sufficient statistics derived from aggregation can be summarized by a number of macro state variables which summarize the effects of past decisions and current information but with a smaller dimension and less requirements on disaggregate details.

In the first application, we explore the micro foundation of the aggregate consumption function by providing a first-order decomposition of aggregate consumption in terms of micro consumer characters. Our result can be interpreted as a generalized version of Auclert's (2019) work. His paper decomposes the aggregate consumption effect of a transitory monetary shock in an economy with a fixed supply of capital into a contribution from a variety of aggregate and redistribution channels. Using our method, we provide a first-order decomposition of aggregate consumption response in an economy allowing for varying capital investment and persistent shocks. All the channels shown in Auclert's decomposition remain while a new redistribution channel—the asset position adjustment (APA) channel—is identified. This channel captures the impact on consumption through individual household's adjustment of asset position. The magnitude of the APA channel mainly depends on the variability of capital investment and the persistence of shocks. Intuitively, the changes of real value of capital asset holding can be decomposed into three parts: the changes of nominal return rate, of general price level, and of capital asset position. Suppose a shock is transitory, a consumer will only reevaluate the present discount value of capital asset without any adjustment of his or her capital asset position. The APA channel is deactivated in this case. However, as long as capital investment is allowed, this channel kicks in. If a consumer's current capital asset is more than the desired level, this consumer will reduce the asset position and consume more relative to the one with less current asset than the desired.

In the second application, we revisit the classic issue on the macroeconomic consequences of changing permanent income inequality. Over the past few decades, the income inequality in the U.S. has been rising rapidly. In particular, a significant share of risking income inequality is accounted by the fixed-effect component of labor income. This component usually captures the returns to skill or ability, which can be interpreted as the permanent income. Many well-known macroeconomic models predict that shifts in the distribution of permanent income are entirely or approximately neutral.[2] Macroeconomic aggregates, such as consumption and interest rate, are independent of permanent income inequality, since consumption is a linear function of permanent income. To break down this linearity property, the role of non-homothetic preference has been emphasized in the literature. Although aggregate consumption is not a linear function in permanent income any more in the presence of non-homothetic preference, we show that the first-order macro impact on aggregate consumption response can be still neutral to any shifts in the distribution of permanent income. However, the impact of varying permanent-income inequality can show up at macro second-order. We provide an explicit characterization on the first- and second-order macro impact of aggregate consumption to a shift of permanent-income inequality, which we refer to as the inequality multiplier. This multiplier is shaped by the microeconomic details of household's preferences and distribution of permanent income.

The remainder of the paper proceeds as follows. Section 3.2 provides an illustrative example to show the key idea of the method and related economic lessons. Section 3.3 describes the method for a class of recursive, disaggregate stochastic dynamic models. Section 3.4 and 3.5 provide applications on how the micro heterogeneity of asset position and permanent income inequality affect aggregate consumption responses. Section 3.6 concludes.

3.2 An Illustrative Example

In this section, we start with an illustrative example of production allowing for sectoral or locational heterogeneity to investigate i) whether there exists a stand-in model that links a specific list of macro variables to micro shocks but not the micro variables; ii)

[2]The macro models adhering to the permanent income hypothesis yield the neutrality result of a linear consumption function in permanent income. Even canonical precautionary-savings models, which are widely known to generate concave consumption functions in current income, also predict a linear consumption function in permanent income. See Straub (2018).

if yes, how to find the first-order approximate stand-in model; iii) economic lessons about the macro impact of micro heterogeneity; iv) what sufficient statistics for micro shocks are; and v) whether it is possible to find higher-order approximate stand-in models and how.

3.2.1 Model Setup

Suppose that there is an arbitrary number of firms indexed by $j \in \mathcal{J}$, where $\mathcal{J}$ can be a finite number or a continuum. The index j can be interpreted as sectors or locations or others. Each firm uses capital and labor to produce output having a function of the form

$$y_j = z_j f(k_j, n_j, j) \tag{3.1}$$

where y_j is output, k_j is capital, n_j is labor and z_j is productivity. The production function f is assumed to be differentiable and constant return to scale but may differ across firms. Final goods are produced by an aggregator over $\{y_j\}$ as follows

$$Y = D(\{y_j\}_{j\in\mathcal{J}}). \tag{3.2}$$

Normalize the price of Y to be one, and then the relative price of y_j is given by $p_j = \frac{\partial D}{\partial y_j}$.

To keep things simple, suppose further that there is an exogenous level of demand d_j for each good j,[3] so that the equilibrium for good j is

$$y_j = d_j. \tag{3.3}$$

The factors, capital and labor, are freely mobile across firms and have prices Q and

[3] This assumption can be relaxed to be endogenous demand but only complicates the expressions.

W respectively. Cost minimization requires that, for each good j, the factor demands

$$\frac{Q}{W} = \frac{f_k(k_j, n_j, j)}{f_n(k_j, n_j, j)}. \tag{3.4}$$

Assume that there is an exogenous total endowment (supply) of capital K^S and labor N^S, so that factor market clearing requires that the demands satisfy

$$K^S = \sum_j k_j, \tag{3.5}$$

$$N^S = \sum_j n_j. \tag{3.6}$$

While this setup is very simple, it can be used to illustrate the questions that a researcher would naturally have in investigating models with multiple sectors or locations of economic activity and, more generally, in models with micro heterogeneity. The first question is whether there exists a *stand-in* model that links a specified list of macro variables (e.g., Y, N, and K in this model) to micro shocks (e.g., productivity shocks $\{z_j\}$), but not the micro variables (e.g., $\{k_j\}$ and $\{n_j\}$), which is generic in the sense that its existence does not depend on the chosen parameters.

In this particular model, the above question is equivalent to ask whether there exists an aggregate production function—a consistent aggregation of micro production functions—that is capable of characterizing the equilibrium behavior of the macro variables. This issue has been intensively investigated by the early literature on aggregation in production functions (e.g., Leontief, 1947; Nataf, 1948; Gorman, 1953; Fisher, 1969). However, the answer is somewhat negative in the sense that the assumptions that make aggregation possible are very stringent and hardly expected to be approximated in actual economies. Take Gorman's (1953) aggregation conditions as an example. He shows that, if the marginal rates of substitution between two types of inputs are the same for all firms, then a necessary and sufficient condition for the consistent aggregate of micro production functions to the aggregate production function $Y = F(K, N)$ is possible if the expansion paths for all firms at a given set of input prices are parallel straight lines through their origins. In order to make sure the expansion paths for all firms are parallel straight lines, micro production functions are additively separable in capital and labor, e.g., log-additive Cobb-Douglas. This condition is in fact an extremely restrictive condition for intersectoral or even interfirm aggregation (Felipe and Fisher, 2003). Therefore, an exact aggregation in production functions is very difficult to reach in general.

3.2.2 The First-order Approximate Stand-in Model

Though an exact stand-in model may be impossible, it is still plausible to have a first-order approximate stand-in model, since micro production functions are log-linear to the first-order approximation and thus they satisfy Gorman's aggregation conditions. In this subsection, we show how to find the first-order approximate stand-in model.

Assume that there is a unique steady state and thus denote the variables with a superscript "$\star$" as the corresponding steady state values. Log-linearizing of the equilibrium conditions for each good, this yields

$$\begin{aligned} \hat{y}_j &= \hat{z}_j + \alpha_j \hat{k}_j + (1 - \alpha_j)\hat{n}_j, \\ \hat{y}_j &= \hat{d}_j, \\ \hat{Q} - \hat{W} &= \xi_j[\hat{n}_j - \hat{k}_j], \end{aligned}$$

where $\alpha_j \equiv \frac{\partial f_j / f_j^\star}{\partial k_j / k_j^\star} = \frac{Q^\star k_j^\star}{p_j^\star y_j^\star}$ is the capital share at the steady state and ξ_j is the reciprocal of the elasticity of substitution between n_j and k_j.[4]

[4]Notably, since the micro production functions are constant return to scale, this yields $\frac{Q}{W} =$

Also log-linearizing the aggregation conditions around steady state, this yields

$$\hat{Y} = \sum_j \theta^y_j \hat{y}_{jt},$$
$$\hat{K} = \sum_j \theta^k_j \hat{k}_{jt},$$
$$\hat{N} = \sum_j \theta^n_j \hat{n}_{jt},$$

where $\theta^y_j = \frac{\partial D/D^\star}{\partial y_j/y^\star_j} = \frac{p^\star_j y^\star_j}{Y^\star}$ is the Domar weight for intermediate good j at steady state. This also implies that $\theta^k_j = \frac{k^\star_j}{K^\star} = \frac{p^\star_j y^\star_j}{Y^\star} \cdot \frac{Qk^\star_j}{p^\star_j y^\star_j} \cdot \frac{Y^\star}{Q^\star K^\star} = \theta^y_j \frac{\alpha_j}{\overline{\alpha}}$ where $\overline{\alpha} = \sum_j \theta^y_j \alpha_j$ denotes the aggregate capital share. Similarly, $\theta^n_j = \frac{n^\star_j}{N^\star} = \theta^y_j \frac{(1-\alpha_j)}{(1-\overline{\alpha})}$.

3.2.2.1 Implications for Aggregate Production

These expressions make it easy to obtain the approximate behavior of aggregate output $\hat{Y} = \sum_j \theta^y_j \hat{y}_{jt}$ with the following representation,

$$\hat{Y} = \hat{A} + \overline{\alpha}\hat{K} + (1-\overline{\alpha})\hat{N}, \tag{3.7}$$

where $\hat{A} \equiv \sum_j \theta^y_j \hat{z}_{jt}$ denotes aggregate productivity. That is, to the first order, the implications of the micro model with sectoral or locational heterogeneity are the same as those of a representative-firm framework, with economy-wide factor shares $\overline{\alpha}$ and $1-\overline{\alpha}$ and aggregate productivity $\hat{A}$ playing the key role. This result is consistent with the classic findings of Solow (1957) and Hulten (1978). As long as a researchre has the knowledge about aggregate factor shares and aggregate productivity but without knowing all details of micro heterogeneity, it is sufficient to characterize the behavior of production and factor input to the aggregate level.

$\frac{f_k(k_j,n_j,j)}{f_n(k_j,n_j,j)} = \Phi(\frac{k_j}{n_j}, j)$.

3.2.2.2 Implications for Aggregate Factor Demand

A researcher may also ask a relevant question: given that the first-order aggregate production function (3.7) holds in general, is it legitimate to assume a representative firm with optimizing behavior under an aggregate production function of Cobb-Douglas form and then make inferences on the first-order macro impacts? The answer is negative. Although the aggregate production function generally holds to the first-order, the factor demand implications are more subtle. Even if there exists a unique inverse relationship between factor intensity and the rates of factor return for all micro production functions to the first order, i.e., $\hat{Q} - \hat{W} = \xi_j[\hat{n}_j - \hat{k}_j]$ for all j, there may not be a unique inverse relationship at the aggregate level, i.e., $\hat{Q} - \hat{W} = \overline{\xi}[\hat{N} - \hat{K}]$ does not hold.

To demonstrate this point, let's first consider the factor demands at micro level. Conditional on market prices of factors and micro level productivity and demand shocks, the micro-level factor demands satisfy

$$\begin{aligned}\hat{n}_j &= \frac{\alpha_j}{\xi_j}(\hat{Q} - \hat{W}) + (\hat{d}_j - \hat{z}_j),\\ \hat{k}_j &= -\frac{(1-\alpha_j)}{\xi_j}(\hat{Q} - \hat{W}) + (\hat{d}_j - \hat{z}_j).\end{aligned}$$

Therefore, this yields the aggregate factor demands

$$\begin{aligned}\hat{N} &= \frac{b}{1-\overline{\alpha}}(\hat{Q} - \hat{W}) + \frac{1}{1-\overline{\alpha}}\sum_j \theta_j^y (1-\alpha_j)(\hat{d}_j - \hat{z}_j),\\ \hat{K} &= -\frac{b}{\overline{\alpha}}(\hat{Q} - \hat{W}) + \frac{1}{\overline{\alpha}}\sum_j \theta_j^y \alpha_j (\hat{d}_j - \hat{z}_j),\end{aligned}$$

where $b = \sum_j \theta_j^y \frac{(1-\alpha_j)\alpha_j}{\xi_j}$.

Suppose that all the micro production functions $f(\cdot)$ are the same, the aggregate factor demands specialize to

$$\hat{N} = \frac{\alpha}{\xi}(\hat{Q} - \hat{W}) + (\hat{Y} - \hat{A}),$$
$$\hat{K} = -\frac{1-\alpha}{\xi}(\hat{Q} - \hat{W}) + (\hat{Y} - \hat{A}).$$

This directly implies a unique inverse relationship at the aggregate level

$$\hat{N} - \hat{K} = \frac{1}{\xi}(\hat{Q} - \hat{W}).$$

In this scenario, if a researcher assumes a representative firm with a Cobb-Douglas aggregate production function with factor shares $\overline{\alpha}$ and $1-\overline{\alpha}$ and aggregate productivity A. The factor demand as an optimizing behavior of the representative firm will be consistent with the above inverse relationship at the aggregate level.

However, in the presence of micro heterogeneity on production functions, the general expression of the relationship between factor intensity and relative return rate of factors is given by

$$\hat{N} - \hat{K} = \frac{b}{\overline{\alpha}(1-\overline{\alpha})}(\hat{Q} - \hat{W}) + \frac{1}{\overline{\alpha}(1-\overline{\alpha})}\hat{X}, \tag{3.8}$$

with the difference

$$\hat{X} = \sum_j \theta_j^y(\overline{\alpha} - \alpha_j)(\hat{d}_j - \hat{z}_j),$$

where $\frac{b}{\overline{\alpha}(1-\overline{\alpha})}$ denotes the aggregate elasticity of substitution between labor and capital, and the difference captures the interaction of fixed heterogeneity in the micro production structures and heterogeneity in the micro shocks. Therefore, at the aggregate level, the factor intensity depends not only on the rates of factor return inversely but also on some other shocks, e.g., $\hat{X}_t$. Thus it is still possible that, in equilibrium, the factor intensity $\hat{N} - \hat{K}$ and the relative rate of factor return $\hat{Q} - \hat{W}$ can both increase, simply because of a positive aggregate factor demand shock $\hat{X} > 0$. This point was also mentioned by Wicksell (1893) and restated by Robinson (1953) which kicked off the famous Cambridge-Cambridge controversy.

3.2.2.3 The First-order Stand-in Model and Sufficient Statistics

The above analysis suggests that, allowing for rich heterogeneity in micro production structures, a researcher can make use of micro equilibrium conditions to derive a stand-in macro model that links the behavior of macro variables to a set of aggregate shocks which potentially are weighted averages of micro shocks. This subsection provides two lessons. From the view of production functions, although micro production functions are not additively separable, an aggregate production function can be somewhat robustly derived to the first-order. But from the view of factor demands, even if the law of diminishing marginal product holds for all micro production functions, there may not exist a unique inverse relationship between factor intensity and the relative rate of factor return at aggregate level. Therefore, to study the macro impact of micro shocks, a research can establish such a stand-in model and gather the information of aggregate elasticities (e.g., factor shares at the aggregate level) and aggregate shocks (e.g., aggregate productivity shocks $\hat{A}$ and aggregate factor demand shocks $\hat{X}$) as sufficient statistics to make macro inferences.

3.2.3 The Second-order Approximate Stand-in Model

3.2.3.1 Standard Second-order Taylor Approximation

Owing to the linearity of first-order approximation, it is relatively easy to find the first-order approximate stand-in model. But aggregation can be difficult to the second-and higher-order approximation. This is because, under standard Taylor expansion, the higher-order systems are no long linear and thus aggregation cannot be applied in general. To see why, let us take the second-order approximation as an example to illustrate the difficulty.

Denote $\Delta y_j^{(s)} = (y_j^{(s)} - y_j^\star)/y_j^\star$ as the log deviation of firm j's output from its steady state to the second-order approximation, where the superscript (s) denotes the second-order approximate variables. The second-order approximation of firm j's micro production function is given by

$$\begin{aligned}\Delta y_j^{(s)} &= \Delta z_j^{(s)} + \alpha_j \Delta k_j^{(s)} + (1-\alpha_j)\Delta n_j^{(s)} \\ &\quad + \frac{1}{2}\varphi_{kk}(\Delta k_j^{(s)})^2 + \frac{1}{2}\varphi_{nn}(\Delta n_j^{(s)})^2 + \varphi_{kn}(\Delta k_j^{(s)})(\Delta n_j^{(s)}),\end{aligned}$$

where $\varphi_{kk} = \frac{\partial \varphi^2}{\partial \tilde{k}_j^2}$ is the second-order derivative of $\varphi(\tilde{k}_j, \tilde{n}_j, j) \equiv f(\exp(\tilde{k}_j), \exp(\tilde{n}_j), j)$ with respect to $\tilde{k}_j = \log k_j$ around steady state and similarly for φ_{nn} and φ_{kn}.

The second-order approximations of the aggregate conditions are given by

$$\begin{aligned}\Delta Y^{(s)} &= \sum_j \theta_j^y \Delta y_j^{(s)} + \frac{1}{2}\sum_j \sum_l \delta_{y_j y_l}(\Delta y_j^{(s)})(\Delta y_l^{(s)}), \\ \Delta K^{(s)} &= \sum_j \theta_j^k \Delta k_j^{(s)} + \frac{1}{2}\sum_j \sum_l \delta_{k_j k_l}(\Delta k_j^{(s)})(\Delta k_l^{(s)}), \\ \Delta N^{(s)} &= \sum_j \theta_j^n \Delta n_j^{(s)} + \frac{1}{2}\sum_j \sum_l \delta_{n_j n_l}(\Delta n_j^{(s)})(\Delta n_l^{(s)}),\end{aligned}$$

where $\delta_{y_j y_l} = \frac{\partial \delta^2}{\partial \tilde{y}_j \partial \tilde{y}_l}$ is the second-order cross derivative of $\delta(\{\tilde{y}_j\}) \equiv D(\{\exp(\tilde{y}_j)\})$ with respect to $\tilde{y}_j = \log y_j$ around steady state and similarly for $\delta_{k_j k_l}$ and $\delta_{n_j n_l}$.

In the second-order approximations, the micro variables are quadratic function and thus no longer additively separable. The Gorman's aggregation conditions in general fail and therefore one cannot apply the method in the previous subsection to construct a stand-in model that links a list of macro variables to micro shocks up to the second-order.

3.2.3.2 Second-order Approximation via Differentials

In this subsection, we show a technique to decompose the second-order terms in a linear manner. Owing to the linearity under this decomposition, we can apply linear aggregation in the second-order approximate system tractably. The technique to decompose the second-order terms borrows from the literature of solving nonlinear rational expectations models. Loosely speaking, the second-order difference of micro variables $\Delta y_j^{(s)}$ can be decomposed as its first- and second-order components, $\Delta y_j^{(s)} = d\tilde{y}_j + \frac{1}{2}d^2\tilde{y}_j$, either in the notion of differentials (Johnston, King, and Lie, 2014) or in the notion of pruning method (Kim, Kim, Schaumburg, and Sims, 2008; Andreasen et al., 2018). The following analysis employs the notion of differentials to decompose the second-order terms and then construct a stand-in model to the second-order.

Taking the second-order differentials of the equilibrium conditions for each good j, this yields

$$d^2\tilde{y}_j = d^2\tilde{z}_j + \alpha_j d^2\tilde{k}_j + (1-\alpha_j)d^2\tilde{n}_j + \frac{1}{2}\varphi_{kk}(d\tilde{k}_j)^2 + \frac{1}{2}\varphi_{nn}(d\tilde{n}_j)^2 + \varphi_{kn}(d\tilde{k}_j)(d\tilde{n}_j),$$

$$d^2\tilde{y}_j = d^2\tilde{d}_j,$$

$$d^2\tilde{Q} - d^2\tilde{W} = \xi_j[d^2\tilde{n}_j - d^2\tilde{k}_j] + \frac{1}{2}\phi_{kk}(d\tilde{k}_j)^2 + \frac{1}{2}\phi_{nn}(d\tilde{n}_j)^2 + \phi_{kn}(d\tilde{k}_j)(d\tilde{n}_j),$$

where $\phi_{kk} = \frac{\partial^2\phi}{\partial\tilde{k}_j^2}$ is the second-order cross derivative of $\phi(\exp(\tilde{k}_j), \exp(\tilde{n}_j)) \equiv \Phi(\frac{k_j}{n_j}, j)$ with respect to $\tilde{k}_j$ around steady state and similarly for ϕ_{nn} and ϕ_{nk}.

The second-order differentials of the aggregation conditions around steady state are

$$d^2\tilde{Y} = \sum_j \theta_j^y d^2\tilde{y}_j + \frac{1}{2}\sum_j\sum_l \delta_{y_j y_l}(d\tilde{y}_j)^2,$$
$$d^2\tilde{K} = \sum_j \theta_j^k d^2\tilde{k}_j + \frac{1}{2}\sum_j\sum_l \delta_{k_j k_l}(d\tilde{k}_j)^2,$$
$$d^2\tilde{N} = \sum_j \theta_j^n d^2\tilde{n}_j + \frac{1}{2}\sum_j\sum_l \delta_{n_j n_l}(d\tilde{n}_j)^2,$$

and the second-order approximate macro variables are given by

$$\Delta Y^{(s)} = d\tilde{Y} + \frac{1}{2}d^2\tilde{Y},$$
$$\Delta K^{(s)} = d\tilde{K} + \frac{1}{2}d^2\tilde{K},$$
$$\Delta N^{(s)} = d\tilde{N} + \frac{1}{2}d^2\tilde{N},$$

Notably, all of the above first-order components (e.g., $d\tilde{y}_j$, $d\tilde{k}_j$, and others) can be sufficiently solved by the first-order approximate equilibrium system as in Subsection 3.2.2. Therefore, when it comes to the second-order approximate system, the first-order components can be treated as given. Therefore, the second-order components of the micro equilibrium conditions are

$$d^2\tilde{y}_j = d^2\tilde{z}_j + \omega_{y_j}^{(s)} + \alpha_j d^2\tilde{k}_j + (1-\alpha_j)d^2\tilde{n}_j,$$
$$d^2\tilde{y}_j = d^2\tilde{d}_j,$$
$$d^2\tilde{Q} - d^2\tilde{W} = \xi_j[d^2\tilde{n}_j - d^2\tilde{k}_j] + \omega_{n_j k_j}^{(s)},$$

where $\omega_{y_j}^{(s)} \equiv \frac{1}{2}\varphi_{kk}(d\tilde{k}_j)^2 + \frac{1}{2}\varphi_{nn}(d\tilde{n}_j)^2 + \varphi_{kn}(d\tilde{k}_j)(d\tilde{n}_j)$, $\omega_{nk,j}^{(s)} \equiv \frac{1}{2}\phi_{kk}(d\tilde{k}_j)^2 + \frac{1}{2}\phi_{nn}(d\tilde{n}_j)^2 + \phi_{kn}(d\tilde{k}_j)(d\tilde{n}_j)$ are denoted as the new second-order micro shocks; and the second-order components of the macro variables are

$$d^2\tilde{Y} = \sum_j \theta_j^y d^2\tilde{y}_j + \omega_Y^{(s)},$$

$$d^2\tilde{K} = \sum_j \theta_j^k d^2\tilde{k}_j + \omega_K^{(s)},$$

$$d^2\tilde{N} = \sum_j \theta_j^n d^2\tilde{n}_j + \omega_N^{(s)},$$

where $\omega_Y^{(s)} \equiv \frac{1}{2}\sum_j \sum_l \delta_{y_j y_l}(d\tilde{y}_j)^2$, $\omega_K^{(s)} = \frac{1}{2}\sum_j \sum_l \delta_{k_j k_l}(d\tilde{k}_j)^2$, and $\omega_N^{(s)} = \frac{1}{2}\sum_j \sum_l \delta_{n_j n_l}(d\tilde{n}_j)^2$ are denoted as the new second-order macro shocks.

3.2.3.3 The Second-order Stand-in Model and Sufficient Statistics

The above decomposition method yields two advantages to construct the second-order stand-in model. On the one hand, the second-order micro and macro shocks are known functions of the cross-product terms of the first-order differentials, which can be treated as given after solving the first-order system. On the other hand, the second-order approximate system has a similar algebraic structure to the first-order approximate system. Therefore, a research can directly apply the way to construct the first-order stand-in model to the second-order one.

Implications for Aggregate Production. Using the second-order relationships between micro and macro output, the second-order differential of aggregate output is given by

$$d^2\tilde{Y} = \overline{\alpha} d^2\tilde{K} + (1-\overline{\alpha})d^2\tilde{N} + [\sum_j \theta_j^y (d^2\tilde{z}_j + \omega_{y_j}^{(s)}) + \omega_Y^{(s)} - \overline{\alpha}\omega_K^{(s)} - (1-\overline{\alpha})\omega_N^{(s)}],$$

This directly implies the second-order approximate behavior of aggregate output

$$\Delta Y^{(s)} = \overline{\alpha}\Delta K^{(s)} + (1-\overline{\alpha})\Delta N^{(s)} + \Delta A^{(s)}, \tag{3.9}$$

with second-order aggregate productivity

$$\Delta A^{(s)} = \sum_j \theta_j^y \Delta z_j^{(s)} + \frac{1}{2}[\sum_j \theta_j^y \omega_{y_j}^{(s)} + \omega_Y^{(s)} - \overline{\alpha}\omega_K^{(s)} - (1-\overline{\alpha})\omega_N^{(s)}].$$

That is, to the second order, the implications of the micro model with sectoral or locational heterogeneity are the same as those of a representative framework but with new interpretations on aggregate productivity shocks. Besides the weighted average of micro productivity shocks, the second-order aggregate productivity shocks also include the dispersion (or cross-product term) of first-order micro variables.

Implications for Aggregate Factor Demand. Similar to the case of first-order approximation, the second-order aggregate factor demand relationship can be represented by

$$\Delta N^{(s)} - \Delta K^{(s)} = \frac{b}{\overline{\alpha}(1-\overline{\alpha})}(\Delta Q^{(s)} - \Delta W^{(s)}) + \Delta X^{(s)}, \tag{3.10}$$

with the second-order aggregate factor demand shocks

$$\Delta X^{(s)} = \frac{1}{\overline{\alpha}(1-\overline{\alpha})}\sum_j \theta_j^y(\overline{\alpha} - \alpha_j)[(\Delta d_j^{(s)} - \Delta z_j^{(s)}) - \frac{1}{2}(\omega_{y_j}^{(s)} - \alpha_j(1-\alpha_j)\omega_{n_j k_j}^{(s)})].$$

Similarly, to the second order, the dispersion (or cross-product terms) of first-order micro responses to micro shocks emerges as a second-order shock.

To sum up, the above analysis suggests that, even allowing for rich heterogeneity in micro production structures, one can make use of micro equilibrium to derive a second-order stand-in model that links the behavior of macro variables to a set of aggregate shocks which are the average and dispersion of micro shocks weighted by their corresponding impact intensities.

3.3 From Micro to Macro: A General Approach

This section describes the general approach in a class of recursive stochastic dynamic systems. We show under what conditions a disaggregate stochastic dynamic equilibrium model can be mapped into a stand-in macro model with a recursive representation. This method also provides a guidance of constructing sufficient statistics to characterize the macro impact of micro heterogeneity. Applications are given in Section 3.4 and 3.5.

3.3.1 Models of Interest

Consider the following class of rational expectations models

$$\mathbb{E}_t \mathcal{F}(\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{e}_{t+1}) = 0, \tag{3.11}$$

where $\mathbf{x}_t = \{x_{jt}\}_{j \in \mathcal{J}}$ denotes a vector of microeconomic unit's decisions and state variables with index j over an arbitrary number of types $\mathcal{J}$, $\mathbf{e}_t = \{e_{jt}\}_{j \in \mathcal{J}}$ represents a vector of microeconomic exogenous variables or shocks across types. $\mathbb{E}_t$ is expectation operator conditional on microeconomic unit's information set at time t. The vector-valued function $\mathcal{F}$ is possibly, and in general, nonlinear. There are n_x variables in the system, which are contained in the vector $\mathbf{x}_t$. The number of exogenous shocks in the system is n_e and the distribution of the exogenous shocks is unspecified for now. So far the only assumptions required are i) the differentiability of $\mathcal{F}$ and ii) the existence and uniqueness of steady state such that $\mathcal{F}(\mathbf{x}^\star, \mathbf{x}^\star, 0) = 0$.

An Example of a Multi-sector Real Business Cycle Model.

Throughout this subsection, we also provide a multi-sector real business cycle model as an example to display how to apply this method. Consider a horizontal economy with multiple sectors indexed by $j = 1, 2, ..., J$. Each sector uses labor n_{jt} and capital k_{jt} as inputs to produce sectoral goods following $y_{jt} = z_{jt} k_{jt}^{\alpha_j} n_{jt}^{1-\alpha_j}$, where z_{jt} is sectoral productivity and α_j is the capital share in sector j which is allowed to be heterogeneous across sectors. Sectoral stock of capital evolves $k_{j,t+1} = (1-\delta)k_{jt} + i_{jt}$, where δ is the depreciation rate of capital. The final goods production function is a CES aggregator over sectoral goods $y_t = (\Sigma_j y_{jt}^{\frac{\varepsilon-1}{\varepsilon}})^{\frac{\varepsilon}{\varepsilon-1}}$, where ε is the elasticity of substitution across sectoral goods. The final goods can be allocated to household's consumption or sectoral firms' investment. In the demand side, there is a representative household who decides aggregate consumption and labor supply to maximize $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t [u(c_t) - v(n_t)]$. The resource constraint of final goods is given by $y_t = c_t + \sum_j i_{jt}$. Assume that labor is freely mobile across sector and the labor market clears, $n_t = \sum_j n_{jt}$.

Now we display the general structure of approximate-model restrictions and solution forms by approximation order as follows. Then we use the first- and second-order approximate system to illustrate more details about how the method works.

The n-th Order Solution.

Using Talyor series expansion, the approximate solution to the system (3.11) to the n-th order is of the form

$$\mathbf{x}_t \approx \mathbf{x}^{\star}[1 + \mathbf{y}_t^{(1)} + \frac{1}{2}\mathbf{y}_t^{(2)} + ... + \frac{1}{n!}\mathbf{y}_t^{(n)}], \tag{3.12}$$

where $\mathbf{x}^\star$ is the vector of deterministic steady-state values of $\mathbf{x}_t$, satisfying $\mathcal{F}(\mathbf{x}^\star, \mathbf{x}^\star, 0) = 0$, and $\mathbf{y}_t^{(i)}$, $i = 1, .., n$, denotes the i-th order element in logarithmic form of the solution. Our approach follows Johnston, King, and Lie (2014) to solve $\mathbf{y}_t^{(i)}$ successively, starting from $i = 1$. $\mathbf{y}_t^{(i)}$ is in fact the solution to the restrictions imposed by the i-th differential of the system (3.11).

General Forms of Approximate-model Restrictions.

The restrictions for the n-th order approximation solution can be shown as a linear system and given by

$$\mathbf{A}\mathbb{E}_t\mathbf{y}_{t+1}^{(i)} = \mathbf{B}\mathbf{y}_t^{(i)} + \mathbf{C}^{(i)}\mathbb{E}_t\mathbf{z}_{t+1}^{(i)}, \tag{3.13}$$

with the driving process evolves according to

$$\mathbf{z}_t^{(i)} = \gamma^{(i)} + \mathbf{D}^{(i)}\mathbf{z}_{t-1}^{(i)} + \mathbf{G}^{(i)}\mathbf{v}_t^{(i)}, \tag{3.14}$$

where the driving variables $\mathbf{z}_t^{(i)}$ are treated as exogenous for each i, but are not necessary exogenous for the whole system (3.11). $\mathbf{A}$ and $\mathbf{B}$ are $n_x \times n_x$ matrices associated with the differentials of the system at steady state. Notably, the matrices $\mathbf{A}$ and $\mathbf{B}$ have no superscript i and hence, are identical for all i. But the matrices $\mathbf{C}^{(i)}$, $\mathbf{D}^{(i)}$, and $\mathbf{G}^{(i)}$ and the vector $\gamma^{(i)}$ are i-th order-specific. The vector of innovations, $\mathbf{v}_t^{(i)}$, has zero mean and a variance-covariance matrix.

Constructing Restrictions for n-th Order Stand-in Macro System.

Now since the system of restrictions is linear, this property provides huge convenience for aggregation. Denote $\mathbf{Y}_t^{(i)}$ as the macro variables to n-th order, which is a $n_Y \times 1$ vector and $\mathbf{H}_y$ as an $n_Y \times n_x$ linear aggregator. For instance, in the illustrative example, the first-order aggregate labor can be shown by $\hat{n}_t = \sum_j \frac{n_j^\star}{n^\star}\hat{n}_{jt}$. Using aggregation relationship among variables $\mathbf{Y}_t^{(i)} = \mathbf{H}_y \mathbf{y}_t^{(i)}$, this yields the n-th order stand-in macro dynamic system

$$\tilde{\mathbf{A}}\mathbb{E}_t \mathbf{Y}_{t+1}^{(i)} = \tilde{\mathbf{B}}\mathbf{Y}_t^{(i)} + \tilde{\mathbf{C}}^{(i)}\mathbb{E}_t \mathbf{Z}_{t+1}^{(i)}, \tag{3.15}$$

where the aggregate driving variables $\mathbf{Z}_t^{(i)}$ are some linear combinations over micro driving variables to n-th order, such that $\mathbf{Z}_t^{(i)} = [\mathbf{i}_z^{(i)}]'\mathbf{z}_t^{(i)}$. $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{C}}^{(i)}$ are matrices associated with the differentials of the system at steady state. The restrictions above are familiar from the literature on linear approximation literature. The solution to $\mathbf{Y}_t^{(i)}$ can be easily obtained and be expressed in a linear state-space form.

Similarly, the n-th order micro dynamics can be shown by the following form

$$\breve{A}_j \mathbb{E}_t \begin{bmatrix} y_{j,t+1}^{(i)} \\ \mathbf{Y}_{t+1}^{(i)} \end{bmatrix} = \breve{B}_j \begin{bmatrix} y_{jt}^{(i)} \\ \mathbf{Y}_t^{(i)} \end{bmatrix} + \breve{C}_j^{(i)} \mathbb{E}_t \begin{bmatrix} z_{j,t+1}^{(i)} \\ \mathbf{Z}_{t+1}^{(i)} \end{bmatrix}, \tag{3.16}$$

where $\breve{A}_j$, $\breve{B}_j$, and $\breve{C}_j^{(i)}$ are matrices associated with the differentials of the system at steady state. Knowing the solution to the n-th order macro impacts, it is easy to plug in the solution to $\mathbf{Y}_t^{(i)}$ into the n-th order micro system and thus, to solve $y_{jt}^{(i)}$ for each j. The solution to $y_{jt}^{(i)}$ can be expressed in a linear state-space form too.

General Forms of Solution to the Stand-in Model.

Given the above model restrictions, the solution to n-th order macro variables can be written in a linear state-space form,

$$\begin{aligned} \mathbf{Y}_t^{(i)} &= \theta_Y^{(i)} + \theta_{YK}^{(i)} \mathbf{K}_t^{(i)}, \\ \mathbf{K}_t^{(i)} &= \theta_K^{(i)} + \theta_{KK}^{(i)} \mathbf{K}_{t-1}^{(i)} + \theta_{KV}^{(i)} \mathbf{V}_t^{(i)}, \end{aligned}$$

for aggregate states $\mathbf{K}_t^{(i)}$ (including endogenous and exogenous state variables). The solution to n-th order micro variables is given by a linear state-space representation,

$$
\begin{aligned}
y_{jt}^{(i)} &= \theta_{y,j}^{(i)} + \theta_{yk,j}^{(i)} k_{jt}^{(i)} + \theta_{YK,j}^{(i)} \mathbf{K}_t^{(i)}, \\
k_{jt}^{(i)} &= \theta_{k,j}^{(i)} + \theta_{kk,j}^{(i)} k_{jt-1}^{(i)} + \theta_{kK,j}^{(i)} \mathbf{K}_{t-1}^{(i)} + \theta_{kv,j}^{(i)} v_{jt}^{(i)},
\end{aligned}
$$

for micro states $k_{jt}^{(i)}$, which also include endogenous and exogenous state variables.

3.3.2 The First-order Stand-in System

Total differentiation of the model $\mathbb{E}_t\mathcal{F}[\exp(\mathbf{y}_{t+1}), \exp(\mathbf{y}_t), \exp(\mathbf{u}_{t+1})] = 0$ yields a set of model restrictions on equilibrium dynamics to the first-order,

$$
\mathbb{E}_t[\mathcal{F}_1 d\mathbf{y}_{t+1} + \mathcal{F}_2 d\mathbf{y}_t + \mathcal{F}_3 d\mathbf{u}_{t+1}] = 0, \tag{3.17}
$$

where the matrices $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$ are the partial derivatives with respect to $\mathbf{y}_{t+1}$, $\mathbf{y}_t$, and $\mathbf{u}_{t+1}$, respectively, evaluated at deterministic steady state. The above system can be solved using standard linear rational expectations solution methods (e.g., King and Watson, 1998; Sims, 2002) in the following form

$$
\mathbf{A}\mathbb{E}_t d\mathbf{y}_{t+1} = \mathbf{B} d\mathbf{y}_t + \mathbf{C}^{(1)}\mathbb{E}_t d\mathbf{u}_{t+1}, \tag{3.18}
$$

where $\mathbf{A} = -\mathcal{F}_1$, $\mathbf{B} = \mathcal{F}_2$, and $\mathbf{C}^{(1)} = \mathcal{F}_3$. In line with this literature, a stable solution exists and is unique if the following conditions are satisfied: (i) there exists a number χ such that $|\mathbf{A}\chi - \mathbf{B}| \neq 0$ and the no-unit-root condition requires that $|\mathbf{A} - \mathbf{B}| \neq 0$; (ii) the relevant generalizations of the Blanchard and Kahn (1980) rank and order conditions must be satisfied.

The standard solution method may subject to a high-dimensionality issue when the model includes a large number of micro heterogeneity. Instead of solving the system (3.18) directly, we propose to construct stand-in macro dynamic systems by order of approximation respectively to reduce dimensionality and then solve these system successively by order or approximation.

In principle, an aggregation can be considered as a linear projection to reduce the rank of the original system. Algebraically speaking, a stand-in macro dynamic system can be constructed if there exists a linear transformation $\mathbf{T}$ such that such that $\mathbf{TA} = \tilde{\mathbf{A}}\mathbf{H}_y$ and $\mathbf{TB} = \tilde{\mathbf{B}}\mathbf{H}_y$. Therefore, aggregate variables are defined by $d\mathbf{Y}_t = \mathbf{H}_y d\mathbf{y}_t$ and the first-order macro system can be shown by

$$\tilde{\mathbf{A}}\mathbb{E}_t d\mathbf{Y}_{t+1} = \tilde{\mathbf{B}} d\mathbf{Y}_t + \mathbb{E}_t \mathbf{U}_{t+1}^{(1)}, \tag{3.19}$$

where $\mathbf{U}_t^{(1)} = \mathbf{T}\mathbf{C}^{(1)} d\mathbf{u}_t$ denotes the first-order macro shocks, which are also a linear combination of micro shocks.[5] Therefore, the solution to the first-order stand-in macro system (3.19) can be written in a recursive, state-space form,

$$\begin{aligned} d\mathbf{Y}_t &= \Pi^{(1)} d\mathbf{K}_t, \\ d\mathbf{K}_{t+1} &= \Phi_K^{(1)} d\mathbf{K}_t + \Phi_U^{(1)} \mathbf{U}_{t+1}^{(1)}, \end{aligned}$$

where $d\mathbf{K}_t$ denotes a vector of macro state variables (including endogenous state variables and exogenous state variables). Similarly, a stable solution exists and is unique if (i) there exists a number χ such that $|\tilde{\mathbf{A}}\chi - \tilde{\mathbf{B}}| \neq 0$ and the no-unit-root condition requires that $|\tilde{\mathbf{A}} - \tilde{\mathbf{B}}| \neq 0$; (ii) the relevant generalizations of the Blanchard and Kahn (1980) rank and order conditions must be satisfied at the aggregate level.

[5] A trivial solution of $\mathbf{T}$ always exists such that $\mathbf{T} = \mathbf{I}$ and thus $\mathbf{A} = \tilde{\mathbf{A}}$, $\mathbf{B} = \tilde{\mathbf{B}}$, $\mathbf{i}_y' = \mathbf{I}$, which means that no aggregation is applied to reduce dimensionality.

The Example of the Multi-section RBC Model.

Denote $\hat{x}_t = d\log x_t$ as the first-order log-approximation around its steady state. Applying the method to the example of multi-sector real business cycle, the first-order stand-in macro system can be constructed through a linear projection such that

$$\begin{aligned}
\phi\hat{n}_t &= \hat{w}_t - \sigma\hat{c}_t, \\
y\hat{y}_t &= c\hat{c}_t + i\hat{i}_t, \\
\hat{k}_{t+1} &= (1-\delta)\hat{k}_t + \delta\hat{i}_t, \\
\hat{c}_t &= \mathbb{E}_t\hat{c}_{t+1} - \sigma^{-1}\frac{r}{R}\mathbb{E}_t\hat{r}_{t+1}, \\
\hat{y}_t &= (\Sigma_j\lambda_j z_{jt}) + \overline{\alpha}\cdot\hat{k}_t + (1-\overline{\alpha})\cdot\hat{n}_t, \\
\mathbb{E}_t\hat{r}_{t+1} &= \mathbb{E}_t\hat{y}_{t+1} - \hat{x}_{t+1}, \\
\hat{x}_{t+1} &= (\varepsilon-1)\cdot(\Sigma_j\lambda_j\mathbb{E}_t\hat{z}_{jt+1}) - \varepsilon[\overline{\alpha} + \frac{1}{\varepsilon}(1-\overline{\alpha})]\cdot\mathbb{E}_t\hat{r}_{t+1} \\
&\quad + \mathbb{E}_t\hat{y}_{t+1} - (\varepsilon-1)(1-\overline{\alpha})\cdot\mathbb{E}_t\hat{w}_{t+1}, \\
\overline{\alpha}\hat{k}_{t+1} &= (\varepsilon-1)\cdot(\Sigma_j\lambda_j\alpha_j\mathbb{E}_t\hat{z}_{jt+1}) - \varepsilon\{\Sigma_j\lambda_j\alpha_j[\alpha_j + \frac{1}{\varepsilon}(1-\alpha_j)]\}\cdot\mathbb{E}_t\hat{r}_{t+1} \\
&\quad + \overline{\alpha}\cdot\mathbb{E}_t\hat{y}_{t+1} - (\varepsilon-1)\{\Sigma_j\lambda_j\alpha_j(1-\alpha_j)\}\cdot\mathbb{E}_t\hat{w}_{t+1},
\end{aligned}$$

where $\hat{x}_{t+1} \equiv \Sigma_j\lambda_j\hat{k}_{jt+1}$ is the augmented aggregate state variable and $\overline{\alpha} \equiv \Sigma_j\lambda_j\alpha_j$. The first-order aggregate system has 2 endogenous and 2 exogenous aggregate state variables. The above first-order aggregate system indicates that there are two endogenous aggregate state variables $\hat{k}_t$ and $\hat{x}_t$ and two exogenous "aggregate shocks", $\sum_j \lambda_j\hat{z}_{jt}$ and $\sum_j \lambda_j\alpha_j\hat{z}_{jt}$.

Intuitively, $\hat{k}_t$ comes from the aggregate production function, $\hat{y}_t = (\Sigma_j\lambda_j z_{jt}) + \overline{\alpha}\cdot\hat{k}_t + (1-\overline{\alpha})\cdot\hat{n}_t$, indicating that a larger $\hat{k}_t$ will lead to a larger aggregate output change $\hat{y}_t$. In terms of aggregate output, what really matters is the mean of capital stock across sectors. The distribution of capital doesn't play a direct role in affecting aggregate output in this model economy.

However, the distribution of capital across sectors affects the capital rental rate, even holding the mean as constant. This is because the equilibrium capital rental rate is pinned down by $\mathbb{E}_t\hat{r}_{t+1} = \mathbb{E}_t\hat{y}_{t+1} - \Sigma_j\lambda_j\hat{k}_{jt+1}$ and $\hat{x}_{t+1} = \sum_j \lambda_j\hat{k}_{jt+1} \neq \hat{k}_{t+1} = \frac{1}{\sum_l \lambda_l\alpha_l}\sum_j \lambda_j\alpha_j\hat{k}_{jt+1}$ if capital shares are not identical across sectors. In fact, the distribution of capital stock across sectors affect the allocative efficiency of input factors, and thus matters for the capital return rate and ultimately influences the consumption path over the dynamics.

Roughly speaking, $\hat{k}_t$ governs the effect of capital on aggregate production, while $\hat{x}_t$ controls the effect of capital on aggregate demand through intertemporal substitution. To get more details, let's discuss two special cases: (i) the case of symmetric capital shares across sectors with $\alpha_j = \alpha$; and (ii) the case of Cobb-Douglas aggregator with $\varepsilon = 1$.

In the case of symmetric capital share across sectors with $\alpha_j = \alpha$, since there is no heterogeneity of capital shares across sectors, we have $\hat{k}_{t+1} = \sum_j \lambda_j\hat{k}_{jt+1} = \hat{x}_{t+1}$. This directly shuts down the effect of capital distribution on the allocative efficiency of input factors. It turns out that only the mean of capital stock across sectors governs both the effects on aggregate production as well as on aggregate demand through capital rental rate. Therefore, the mean of capital turns to be the only sufficient statistic.

Now consider the case of Cobb-Douglas aggregator with $\varepsilon = 1$. This case mimics the Long-Plosser RBC model with heterogeneous output elasticities of input factors. The reason why $\hat{x}_{t+1} = \hat{k}_{t+1}$ is because of the Cobb-Douglas aggregator. Knowing that the capital rental rate follows $\mathbb{E}_t\hat{r}_{t+1} = \frac{1}{\varepsilon}\mathbb{E}_t\hat{y}_{t+1} + (1 - \frac{1}{\varepsilon})\mathbb{E}_t\hat{y}_{jt+1} - \hat{k}_{jt+1}$, the elasticity ε controls the effects of sectoral prices p_{jt}. Under Cobb-Douglas aggregator with $\varepsilon = 1$, the effects of sectoral prices are fully offset by the effects of sectoral output. Therefore, the equilibrium capital rental rate will only depend on the mean of capital stock across sectors, such that $\mathbb{E}_t \hat{r}_{t+1} = \mathbb{E}_t \hat{y}_{t+1} - \frac{1}{\sum_l \lambda_j \alpha_j} \sum_j \lambda_j \alpha_j \hat{k}_{jt+1} = \mathbb{E}_t \hat{y}_{t+1} - \hat{k}_{t+1}$.

Taking into stock, most of the traditional literature assumes either CES aggregator with homogeneous capital shares over intermediate goods production (e.g., New Keynesian models) or Cobb-Douglas aggregator but allowing for heterogeneous production elasticity of input factors (e.g., Long-Plosser RBC models with production network). These types of models behave as a standard RBC model with aggregate productivity shocks $\hat{z}_t \equiv \sum_j \lambda_j \hat{z}_{jt}$. As long as the two assumptions are relaxed simultaneously, micro-level productivity shocks can affect aggregate dynamics differently from the standard RBC model. The macro impacts of micro shocks depend not only on the mean of the shocks but also on its distribution, given heterogeneous micro-level production structures.

First-order Micro Dynamic System.

Extracting type j's micro dynamics from the original system (3.18), this yield the first-order micro dynamic system for type j:

$$\breve{A}_j \mathbb{E}_t \begin{bmatrix} dy_{j,t+1} \\ d\mathbf{Y}_{t+1} \end{bmatrix} = \breve{B}_j \begin{bmatrix} dy_{jt} \\ d\mathbf{Y}_t \end{bmatrix} + \breve{C}_j^{(1)} \mathbb{E}_t \begin{bmatrix} du_{j,t+1} \\ \mathbf{U}_{t+1}^{(1)} \end{bmatrix}, \tag{3.20}$$

Since the first-order dynamics of aggregate variables can be solved as a recursive, state-space system, the solution to the first-order micro system can be expressed as follows,

$$
\begin{aligned}
dy_{jt} &= \theta^{(1)}_{y,j} + \theta^{(1)}_{yk,j} dk_{jt} + \theta^{(1)}_{YK,j} d\mathbf{K}_t, \\
dk_{jt} &= \theta^{(1)}_{k,j} + \theta^{(1)}_{kk,j} dk_{jt-1} + \theta^{(1)}_{kK,j} d\mathbf{K}_{t-1} + \theta^{(1)}_{kv,j} du_{jt},
\end{aligned}
$$

where dk_{jt} denotes a vector of micro state variables which also include endogenous and exogenous state variables.

3.3.3 The Second-order Stand-in System

Total differentiation of the first-order restrictions provides the second-order restrictions on equilibrium dynamics,

$$
\mathbb{E}_t[\mathcal{F}_1 d^2\mathbf{y}_{t+1} + \mathcal{F}_2 d^2\mathbf{y}_t + \mathbf{v}^{(2)}_{t+1}] = 0, \tag{3.21}
$$

where the second-order driving force $\mathbf{v}^{(2)}_{t+1}$ can be expressed as a function of various cross products of the elements of the first-order differentials and driving force,

$$
\begin{aligned}
\mathbf{v}^{(2)}_{t+1} = \mathcal{F}_{11}(d\mathbf{y}_{t+1} \otimes d\mathbf{y}_{t+1}) + 2\mathcal{F}_{12}(d\mathbf{y}_{t+1} \otimes d\mathbf{y}_t) + 2\mathcal{F}_{13}(d\mathbf{y}_{t+1} \otimes d\mathbf{u}_{t+1}) \\
+ \mathcal{F}_{22}(d\mathbf{y}_t \otimes d\mathbf{y}_t) + 2\mathcal{F}_{23}(d\mathbf{y}_t \otimes d\mathbf{u}_{t+1}) + \mathcal{F}_{33}(d\mathbf{u}_{t+1} \otimes d\mathbf{u}_{t+1}).
\end{aligned}
$$

Unlike the first-order driving force $d\mathbf{u}_t$, the second-order driving force is endogenous from the perspective of the whole model. However, as the first-order differentials can be solved by the first-order macro and micro systems, the second-order driving force can be considered as exogenous to the second-order differentials. The above second-order system (3.21) can be also solved using standard linear rational expectations solution methods in the following form,

$$\mathbf{A}\mathbb{E}_t d^2\mathbf{y}_{t+1} = \mathbf{B}d^2\mathbf{y}_t + \mathbb{E}_t\mathbf{v}_{t+1}^{(2)}.$$

Notably, the coefficient matrices $\mathbf{A}$ and $\mathbf{B}$ are the same as in the first-order system (3.18); the only difference comes from the driving process. This property yields a huge advantage on aggregation. The aggregation operator $\mathbf{T}$ in fact can be applied to the second- or higher order approximate stand-in systems. Therefore, the second-order stand-in macro system can be represented by

$$\tilde{\mathbf{A}}\mathbb{E}_t d^2\mathbf{Y}_{t+1} = \tilde{\mathbf{B}}d^2\mathbf{Y}_t + \mathbb{E}_t\mathbf{V}_{t+1}^{(2)}, \tag{3.22}$$

where the second-order aggregate components are defined by $d^2\mathbf{Y}_t = \mathbf{i}_y' d^2\mathbf{y}_t$ and the second-order driving process is given by $\mathbf{V}_t^{(2)} = \mathbf{T}\mathbf{v}_t^{(2)}$. The solution to the second-order macro system () can be written in a recursive, state-space form,

$$d^2\mathbf{Y}_t = \Pi^{(2)}\mathbf{K}_t^{(2)},$$
$$\mathbf{K}_{t+1}^{(2)} = \Phi_K^{(2)}\mathbf{K}_t^{(2)} + \Phi_U^{(2)}\mathbf{V}_{t+1}^{(2)},$$

where $\mathbf{K}_t^{(2)}$ denotes a vector of macro state variables. Since the coefficient matrices to the second-order stand-in macro system are identical to the ones to the first-order stand-in macro system, a stable second-order solution exists and is unique if the required conditions for the first-order stand-in macro system are satisfied.

Similarly, extracted from the second-order dynamic system, the second-order micro dynamic system for type j can be shown as follows,

$$\breve{A}_j \mathbb{E}_t \begin{bmatrix} d^2 y_{j,t+1} \\ d^2 \mathbf{Y}_{t+1} \end{bmatrix} = \breve{B}_j \begin{bmatrix} d^2 y_{jt} \\ d^2 \mathbf{Y}_t \end{bmatrix} + \breve{C}_j^{(1)} \mathbb{E}_t \begin{bmatrix} v_{j,t+1}^{(2)} \\ \mathbf{V}_{t+1}^{(2)} \end{bmatrix}. \tag{3.23}$$

The solution to the second-order micro system can be expressed by

$$\begin{aligned} d^2 y_{jt} &= \theta_{y,j}^{(2)} + \theta_{yk,j}^{(2)} k_{jt}^{(2)} + \theta_{YK,j}^{(2)} \mathbf{K}_t^{(2)}, \\ k_{jt}^{(2)} &= \theta_{k,j}^{(2)} + \theta_{kk,j}^{(2)} k_{jt-1}^{(2)} + \theta_{kK,j}^{(2)} \mathbf{K}_{t-1}^{(2)} + \theta_{kv,j}^{(2)} v_{jt}^{(2)}, \end{aligned}$$

where $k_{jt}^{(2)}$ denotes a vector of micro state variables for second-order.

3.3.4 Discussions on Higher-order Stand-in Systems

It is in fact straightforward to extend to the third- and higher order by following similar steps as the above. To obtain the n-th order differential restrictions, one needs to take the total differentiation to the restrictions in the $(n-1)$-th order. The whole procedure of aggregation and solution are convenient due to the following two advantages. On the one hand, these model restrictions have similar form to those in any lower order, i.e., the coefficient matrices $\mathbf{A}$ and $\mathbf{B}$ are order-invariant. Therefore, the aggregation matrix $\mathbf{T}$ can be repeatedly used for any order of approximation. On the other hand, the n-th order driving process is a function of cross products of lower order differentials and driving process, which are solved in $(n-1)$-th order restrictions. The whole dynamics thus can be solved by order of approximation successively.

3.4 Application I: The Aggregate Consumption Function

This section provides a first-order decomposition of the aggregate consumption function and identify a new redistribution channel—the asset position adjustment (APA) channel—which captures the impact on consumption through the adjustment of asset position. This decomposition can be interpreted as a generalized version of Auclert's (2019) result. The magnitude of the APA channel mainly depends on the variability of capital investment and the persistence of shocks. Our analysis starts with a stylized consumption-saving problem in a partial equilibrium model. We then characterize the first-order responses of both individual and aggregate consumption to shocks.

3.4.1 A Stylized Consumption-Saving Problem

Consider a stylized consumption-saving problem in a dynamic, incomplete-market partial equilibrium setting. A typical consumer i faces idiosyncratic income uncertainty and chooses consumption c_{it} and asset position for next period a_{it+1} to maximize his or her life-time expected utility $\mathbb{E}_0 \sum_j \beta^t u(c_{it})$. Assume that each individual supplies one unit of labor inelastically.[6] The real wages $\{w_{it}\}$ follow a stochastic process. The consumer can only trade risk-free nominal assets subject to some trading frictions. The consumer's budget constraint at date t is given by

$$P_t c_{it} + a_{it+1} = P_t w_{it} + R_t a_{it}, \tag{3.24}$$

and the trading of asset is limited by a generic borrowing constraint

$$\lambda(a_{it+1}) \geq 0,$$

[6]It can be relaxed to be elastic labor supply. The main results still preserve.

where P_t is the general level of prices and R_t is the nominal interest rate.

The individual's consumption Euler equation is $u_c(c_{it})/P_t = \beta \mathbb{E}_t u_c(c_{it+1})/P_{t+1}R_{t+1} + \phi_{it}$, where ϕ_{it} is the Lagrangian multiplier with respect to the borrowing constraint. Because ϕ_{it} can be zero at some states, the standard log-linearization method can not be applied. To make it work, let's define a change of variable to represent the slackness of borrowing constraint, $\theta_{it} \equiv 1 - \frac{\phi_{it}}{u_c(c_{it})/P_t}$, the consumption Euler equation can be represented by

$$\theta_{it} \cdot u_c(c_{it})/P_t = \beta \mathbb{E}_t u_c(c_{it+1})/P_{t+1}R_{t+1}, \tag{3.25}$$

for any individuals and at any states, $\theta_{it} \in (0, 1]$, thus θ_{it} are well-defined in terms of log-deviation at local values. The complement-slackness condition for borrowing constraint is given by

$$(1 - \theta_{it}) \cdot u_c(c_{it})/P_t \cdot \lambda(a_{it+1}) = 0. \tag{3.26}$$

3.4.2 The First-order Response of Individual Consumption

To keep notations compact, let's define $\sigma(c_i) \equiv -\frac{u_{cc}(c_i) \cdot c_i}{u_c(c_i)}$ as the local intertemporal elasticity of substitution and $\lambda_a(a_i) \equiv \frac{\partial \lambda}{\partial a} a_i$ as the local marginal intensity of borrowing constraint. Denote $\hat{c}_{it} = d \log c_{it} = d\tilde{c}_{it}$ as log-deviations of the variables at their local values. The first-order logarithmic approximation of individual's optimality conditions around the local values $(c_i, a_i, \theta_i, w_i, R, P)$ are

$$\begin{aligned}\hat{\theta}_{it} - \sigma(c_i)\cdot\hat{c}_{it} - \hat{P}_t &= -\sigma(c_i)\cdot\mathbb{E}_t\hat{c}_{it+1} - \mathbb{E}_t\hat{P}_{t+1} + \mathbb{E}_t\hat{R}_{t+1}\\ \hat{\theta}_{it} &= \frac{1-\theta_i}{\theta_i}\Big(\lambda_a(a_i)\cdot\hat{a}_{it+1} - \sigma(c_i)\cdot\hat{c}_{it} - \hat{P}_t\Big)\\ Pc_i(\hat{c}_{it}+\hat{P}_t) + a_i\hat{a}_{it+1} &= Pw_i(\hat{w}_{it}+\hat{P}) + Ra_i(\hat{a}_{it}+\hat{R}_t).\end{aligned}$$

To explicitly characterize the consumption responses to a perturbation of the economic environment, let's assume that $\hat{R}_t$, $\hat{P}_t$, and $\hat{w}_{it}$ are exogenous AR(1) process with persistence ρ_r, ρ_p, and ρ_{wi}, respectively. Applying the undetermined-coefficient method to solve this linear rational expectations model, the solution of consumption and asset position for next period are

$$\hat{c}_{it} = \gamma^{\star}_{ca,i}\hat{a}_{it} + \gamma^{\star}_{cw,i}\hat{w}_{it} - \gamma^{\star}_{cP,i}\hat{P}_t + (\gamma^{\star}_{cR1,i} - \gamma^{\star}_{cR2,i})\hat{R}_t \tag{3.27}$$

$$\hat{a}_{it+1} = \gamma^{\star}_{aa,i}\hat{a}_{it} + \gamma^{\star}_{aw,i}\hat{w}_{it} + \gamma^{\star}_{aP,i}\hat{P}_t + \gamma^{\star}_{aR,i}\hat{R}_t \tag{3.28}$$

where $\{\gamma^{\star}_{h,i}\}$ are coefficients associating the local values and structural parameters in the model. The solution to these coefficients are shown in the Appendix.

Before characterizing the first-order response of individual's consumption, let's provide the definition of individual's disposable income and marginal propensities to consume. Denote individual's disposable income in real terms by $y_{it} \equiv w_{it} + R_t a_{it}/P_t$ and the marginal propensity to consume (in dollar-to-dollar sense) as

$$MPC_i \equiv \frac{dc_{it}}{dw_{it}} = \frac{c_i}{w_i}\cdot\gamma^{\star}_{cw,i}.$$

Now the first-order response of individual consumption is shown by the following proposition.

Proposition 1 (First-order Micro Consumption Responses). *The first-order response of individual's consumption in response to $\hat{R}_t$, $\hat{P}_t$ and $\hat{y}_{it}$ is*

$$c_i\hat{c}_{it} = MPC_i\left[y_i\hat{y}_{it} - NNP_i \cdot \hat{P}_t + URE_i \cdot \hat{R}_t + APA_i \cdot a_i\hat{a}_{it}\right] - SUB_i \cdot \hat{R}_t, \quad (3.29)$$

where $MPC_i = \frac{c_i}{w_i}\gamma^\star_{cw,i}$ represents individual i's marginal propensity to consume, $NNP_i = \frac{w_i\gamma^\star_{cP,i}}{\gamma^\star_{cw,i}} - \frac{Ra_i}{P}$ represents the Fisher channel, $URE_i = \frac{w_i\gamma^\star_{cR1,i}}{\gamma^\star_{cw,i}} - \frac{Ra_i}{P}$ represents the unhedged interest rate exposure channel, $APA_i = \frac{w_i\gamma^\star_{ca,i}}{a_i\gamma^\star_{cw,i}} - \frac{R}{P}$ represents the asset position adjustment channel, and $SUB_i = c_i\gamma^\star_{cR2,i}$ represents the substitution channel.

Broadly speaking, the above decomposition can be classified into two effects: the wealth effect and the substitution effect. The substitution effect is associated with the IES and represented by $SUB_i \cdot \hat{R}_t$. The wealth effect is related to several parts: $y_i\hat{y}_{it} + APA_i a_i\hat{a}_{it} - NNP_i\hat{P}_t + URE_i\hat{R}_t$. The first term $y_i\hat{y}_{it}$ is the traditional effect from the change in disposable income.

The second term $NNP_i \cdot \hat{P}_t$, where NNP is short for "net nominal position", represents the effect from an increase in the level of nominal prices. If the consumer is a nominal lender, his or her NNP should be negative, then an increase of nominal price hurts his or her real returns, leading to a reduction on net wealth. Conversely, if the consumer is a nominal borrower, he or she gains when facing an increase of nominal price. This is so called the Fisher channel.

The third term $URE_i \cdot \hat{R}_t$, where URE is short for "unhedged interest rate exposure", characterize the effect from the change in the real interest rate, holding the asset position unchanged. If the consumer is a real lender, his or her URE should be positive, then he or she will benefit from an increase of real interest rate.

The fourth term $APA_i \cdot a_i\hat{a}_{it}$, where APA is short for "asset position adjustment", captures the effect from the change in asset (or capital) position. Intuitively, the changes of real value of asset holding $R_t a_{it}/P_t$ can be decomposed into three parts: $\hat{R}_t + \hat{a}_{it} - \hat{P}_t$. Suppose a shock is transitory, the consumer will only reevaluate the present discount value of asset holding but have no incentive to adjustment his or her asset position. Therefore, in this case, the APA channel is deactivated, which is in line with Auclert's (2019) findings. However, if a shock is persistent, instead of being transitory, then the consumer may have incentive to adjust the asset position. If the consumer's current asset position is positive, it means that he or she has relative more asset holding than the desired level. Therefore, the consumer will reduce the asset position and consume more relative to the one with negative asset position.

3.4.3 The First-order Response of Aggregate Consumption

3.4.3.1 Economic Environment

Consider a closed economy populated by $\mathcal{I}$ heterogeneous types of agents, in which each type i has a unit mass of individuals and each in an idiosyncratic state $s_{it} \in S_i$. I use the notation $\mathbb{E}_{\mathcal{I}}(z_{it})$ for the cross-sectional average of any variable z_{it}, taken over individual types $\mathcal{I}$ and idiosyncratic states S_i. Therefore, aggregate (per capita) consumption C_t is equal to average individual consumption $\mathbb{E}_{\mathcal{I}}(c_{it})$, and thus, the log-deviation of aggregate consumption has the relationship to the one of individual consumption as follows $C\hat{C}_t = \mathbb{E}_{\mathcal{I}}(c_i \hat{c}_{it})$.

Consumers. Each consumer type i in state s_{it} has a stochastic endowment of $e_i(s_{it})$ efficient units of work, and thus receives a wage of $w_{it} = w_t \cdot e_i(s_{it})$, where w_t is the real wage per efficient unit of labor. The agent's gross-of-tax income is $y_{it} = w_{it} + R_t a_{it}/P_t$. The economy has a varying stock of aggregate capital K_t with a depreciation rate $\delta \in (0, 1)$. The aggregate amount of asset equals to the stock of aggregate capital

$P_t K_t = \mathbb{E}_{\mathcal{I}}(a_{it})$. The stock of aggregate capital evolves as $K_{t+1} = (1-\delta)K_t + I_t$. Consumers choose consumption c_{it} and asset holding a_{it+1} to maximize their life-time expected utility $\mathbb{E}_0 \sum_t \beta^t u(c_{it})$. Individual's trading on asset holding is limited by a generic borrowing constraint, $\lambda(a_{it+1}) \geq 0$.

Firms and Production. Assume there is a competitive firm that produces a final good by using capital and labor following a production function: $Y_t = A_t F(K_t, N_t)$, where $N_t = \mathbb{E}_{\mathcal{I}}(e_{it})$ and A_t are the aggregate amount of labor and aggregate productivity level. Assume A_t following a stochastic process. In competitive factor markets, the market wage rate equals to marginal product of labor and the capital rental rate equals to marginal product of capital.

Government and Market Clearing. A government issues nominal short-term debt B_t and levies lump-sum tax to finance its expenditure G_t. Thus, its nominal budget constraint is $Q_t B_{t+1} = P_t G_t + B_t - P_t \mathbb{E}_{\mathcal{I}}(t_{it})$, where $Q_t = \frac{1}{R_t}\frac{P_t}{P_{t+1}}$ is the one-period nominal discount rate and $P_t t_{it}$ are lump-sum tax for type i agents. In equilibrium, the markets for capital, labor and final goods all clear. The resource constraint of final goods is $Y_t = C_t + I_t + G_t$. Under the assumptions made here, all consumers in this economy essentially solve the problem in section 4.1.

3.4.3.2 First-order Response of Aggregate Consumption

We are interested in the first-order aggregate consumption response to a perturbation of this environment in which individual gross incomes $\hat{y}_{it}$, nominal prices $\hat{P}_t$, and the interest rate $\hat{R}_t$ change at $t = 0$. The shocks are unexpected but can be transitory or persistent. The decomposition of the first-order aggregate consumption response is characterized by the following proposition.

Proposition 2 (First-order Aggregate Consumption Responses). *The first-order response of aggregate consumption in response to $\hat{R}_t$, $\hat{P}_t$, $\hat{Y}_t$ and $\hat{y}_{it}$ is*

$$\begin{aligned} C\hat{C}_t = & \underbrace{\mathbb{E}_{\mathcal{I}}\left[MPC_i\right]Y\hat{Y}_t}_{\text{Aggregate income channel}} + \underbrace{Cov_{\mathcal{I}}\Big(MPC_i,(y_i\hat{y}_{it}-Y\hat{Y}_t)\Big)}_{\text{Earnings heterogeneity channel}} \\ & + \underbrace{\mathbb{E}_{\mathcal{I}}\left[MPC_i\cdot APA_i\right]K\hat{K}_t}_{\text{Aggregate capital channel}} + \underbrace{Cov_{\mathcal{I}}\Big(MPC_i\cdot APA_i,(a_i\hat{a}_{it}-K\hat{K}_t)\Big)}_{\text{Asset position heterogeneity channel}} \\ & - \underbrace{Cov_{\mathcal{I}}\left(MPC_i,NNP_i\right)\hat{P}_t}_{\text{Fisher channel}} + \underbrace{Cov_{\mathcal{I}}\left(MPC_i,URE_i\right)\hat{R}_t}_{\text{Interest rate exposure channel}} - \underbrace{\mathbb{E}_{\mathcal{I}}\left[SUB_i\right]\hat{R}_t}_{\text{Substitution channel}}. \end{aligned} \tag{3.30}$$

This proposition shows that the first-order response of aggregate consumption to a macroeconomic shock can be decomposed into several channels and these channels can be characterized by a small set of sufficient statistics.

To understand these channels better, we first analyze the channels that show up in a textbook-style representative agent model ($\mathcal{I}=1$). In this case, there is no earning heterogeneity, no asset position heterogeneity, no net nominal position heterogeneity, and no heterogeneous exposure of unhedged interest rate. Therefore, the first-order response of aggregate consumption is given by

$$C\hat{C}_t = MPC\cdot Y\hat{Y}_t + MPC\cdot APA\cdot K\hat{K}_t - SUB\cdot\hat{R}_t$$

The first term $MPC\cdot Y\hat{Y}_t$ stands for income effect, the second term $MPC\cdot APA\cdot K\hat{K}_t$ represents the effect from capital stock, and the third term $SUB\cdot\hat{R}_t$ is a substitution effect. If aggregate capital is fixed, then the aggregate capital channel is deactivated. In contrast, if aggregate capital is varying, then aggregate consumption responses are amplified through the aggregate capital channel. If aggregate capital stock is higher than its steady state level, the representative consumer will reduce capital stock for next period and thus consume more.

Away from this benchmark, now consider the case with different marginal propensities to consume. $Cov_{\mathcal{I}}\left(MPC_i, NNP_i\right)$ represents the Fisher channel. Usually, net nominal borrowers have higher marginal propensities to consume than net nominal lenders. Thus, an increase of price level (i.e., inflation) will increase aggregate consumption relative to the representative-agent benchmark. $Cov_{\mathcal{I}}\left(MPC_i, URE_i\right)$ stands for the unhedged interest rate exposure channel. Since agents with unhedged real borrowing generally have higher MPCs than unhedged savers. Aggregate consumption rises more in response to a decline of real interest rate relative to the representative-agent benchmark. $Cov_{\mathcal{I}}(MPC_i, (y_i \hat{y}_{it} - Y\hat{Y}_t))$ characterizes the earnings heterogeneity channel. Low-income agents generally have high MPCs. Suppose income inequality is dampened in response to a macro shock, this shock can amplify the response of aggregate consumption, and vice verse. Finally, $Cov_{\mathcal{I}}(MPC_i \cdot APA_i, (a_i \hat{a}_{it} - K\hat{K}_t))$ shows the asset position heterogeneity channel. Agents with low asset position tend to have high MPCs and APAs. Suppose asset dispersion is dampened in response to a macro shock, this shock can also amplify the response of aggregate consumption.

3.5 Application II: Inequality Multipliers

Applying our method, this section reevaluates the macro impact of income inequality. The income inequality has been rising in the U.S. over the past few decades. In particular, a significant share of the rising income inequality is accounted by the fixed-effect component of labor income. This component usually captures the returns to skill or ability, which can be interpreted as the permanent component and thus we refer to it as *permanent income* in this section.

Many well-known macroeconomic models, which include a broad set of macro models with homothetic preference over consumption, predict that shifts in the distribution of permanent income are entirely or approximately neutral (Straub, 2018). Macroeconomic aggregates, such as consumption, are independent of permanent income inequality since consumption is a linear function of permanent income.[7] To explore the macro impact of permanent income inequality, we break down this linearity relationship by emphasizing the role of non-homothetic preferences. This section studies a simple macro model with heterogeneous returns to skill, which mimics the distribution of permanent income to analyze both the first- and second-order macro impacts of aggregate consumption responses.

Although aggregate consumption is not a linear function in permanent income in the presence of non-homothetic preference, we show that the first-order macro impact on aggregate consumption response can still be neutral to a shift in permanent income inequality. However, the major impact of permanent income inequality shows up at macro second-order. We provide an explicit characterization on the second-order macro impact of aggregate consumption response to a shift of permanent income inequality, which we refer to as the inequality multiplier.

3.5.1 Basic Structure

Households and Preferences. Consider an economy populated by a continuum of individual households indexed by $i \in [0, 1]$. Each household supplies one unit of labor inelastically but with heterogeneous labor productivity z_i. The labor produc-

[7] The canonical precautionary-savings models (e.g., Aiyagari, 1994; Carroll, 1997), which are widely known to generate concave consumption functions in *current* income, still predict a linear consumption function in permanent income.

tivity is drawn at the initial period, following a cumulative density function $G(z)$, and remains the same for each individual permanently. Households decide their consumption stream to maximize their life-time expected utility $\mathbb{E}_0 \sum_t \beta^t u(c_{it})$ subject to the budget constraint $c_{it} + k_{it+1} = z_i W_t + R_t k_{it}$, where W_t and R_t are market wage rate and capital rental rate. To normalize aggregate labor productivity, assume $\int z_i di = \overline{N} = 1$.

Firms and Production. Assume a representative producing final goods with a Cobb-Douglas production function, $Y_t = A_t K_t^{\alpha} \overline{N}^{1-\alpha}$, where A_t is aggregate productivity level following a stochastic process. The capital evolves as $K_{t+1} = (1-\delta)K_t + I_t$, where δ is the capital depreciation rate. Competitive factor markets imply market wage and capital rental rate as $W_t = (1-\alpha)A_t K_t^{\alpha}$ and $R_t = \alpha A_t K_t^{\alpha-1} + 1 - \delta$. Notably, the heterogeneous labor productivity can be also interpreted as the labor income share of skill types. Assume the final goods production function to be $Y_t = A_t K_t^{\alpha} \Pi_i l^{(1-\alpha)z_i}$, where $l = 1$ is the one unit of labor inelastically supplied by households. Under this specification, the market wage rate for type i worker is $w_{it} = z_i W_t$. In this section, our main focus will be a shift in the distribution of labor income share $\{z_i\}$, which induces a shift in the distribution of skill wages $\{w_{it}\}$, i.e., the permanent income inequality. Assume there is no government activity. The final goods market clears such that $Y_t = C_t + I_t$.

3.5.2 The First-order Macro Impacts: A Neutrality Result

Now let's turn to the first-order macro impacts of permanent income inequality. To keep notations compact, define $\sigma(c_i) \equiv -\frac{u_{cc}(c_i) \cdot c_i}{u_c(c_i)}$ as the inverse of the *micro* intertemporal elasticity of substitution (IES) for household i, where c_i is household i's consumption level at steady state. The first-order micro and macro consumption

Euler equations are shown by the following lemma.

Lemma 1. *To the first-order, the micro and macro consumption Euler equations are, respectively,*

$$\hat{c}_{it} = \mathbb{E}_t \hat{c}_{it+1} - \frac{1}{\sigma(c_i)} \mathbb{E}_t \hat{R}_{t+1}$$

and

$$\hat{C}_t = \mathbb{E}_t \hat{C}_{t+1} - \left(\int \frac{c_i / C^\star}{\sigma(c_i)} di \right) \mathbb{E}_t \hat{R}_{t+1},$$

where $\overline{\gamma} \equiv \int \frac{c_i/C^\star}{\sigma(c_i)} di$ *can be interpreted as the macro intertemporal elasticity of substitution.*

Let's first compare the cases of two commonly-used functional forms: CRRA utility $u(c_{it}) = \frac{c_{it}^{1-\sigma}}{1-\sigma}$ and Stone-Geary utility $u(c_{it}) = \frac{(c_{it} - \underline{c})^{1-\sigma}}{1-\sigma}$. In the CRRA utility case, the inverse of micro IES is constant and identical across households with possibly different income or wealth levels, $\sigma(c_i) = \sigma$. Thus, the macro IES is also a constant, $\int \frac{c_i/C^\star}{\sigma(c_i)} di = \frac{1}{\sigma}$, which is independent of the distribution of household's consumption. Therefore, any shift in the distribution of permanent income plays no role in affecting aggregate consumption responses. This result is expected because a large set of macro models with homothetic preference over consumption predicts that consumption is a linear function of permanent income. In the Stone-Geary utility case, the inverse of micro IES is dependent of household's consumption level as well as his or her permanent income, $\sigma(c_i) = \sigma \frac{c_i}{c_i - \underline{c}}$, which implies that high income households (i.e., high z_i and thus high c_i) have relative low IES and thus small micro consumption responses to an interest rate shock. However, under the Stone-Geary utility specification, the macro IES is still a constant, $\int \frac{c_i/C^\star}{\sigma(c_i)} di = \frac{1}{\sigma}(C^\star - \underline{c})$, which is also independent of the distribution of permanent income. Although consumption is not a linear function of permanent income in the presence of non-homothetic preference, this result indicates that the distribution of permanent income does not affect aggregate consumption responses to first-order neither.

Furthermore, the above neutrality result of the first-order macro impact still holds in case of heterogeneous preferences across households. Suppose that household's utilities over consumption are allowed to be heterogeneous, e.g., $u(c_{it}) = \frac{c_{it}^{1-\sigma_i}}{1-\sigma_i}$ for CRRA utility or $u(c_{it}) = \frac{(c_{it}-\underline{c}_i)^{1-\sigma_i}}{1-\sigma_i}$ for Stone-Geary utility, where $\{\sigma_i, \underline{c}_i\}$ are drawn from a cumulative density function $H(\sigma, \underline{c})$ at the initial period. As long as the distribution of household's heterogeneous utility is independent of the distribution of permanent income across households, the first-order aggregate consumption responses are still neutral to any shift in the distribution of permanent income.

Although the homotheticity of preference does play a role in affecting whether aggregate consumption is linear function in permanent income or not, the neutrality of first-order macro consumption responses actually depends on the micro structure of household's intertemporal preferences. The following proposition summarizes the neutrality result of first-order aggregate consumption responses to a change of permanent income distribution.

Proposition 3. *If individual household's instantaneous utility satisfies the following condition: the inverse of absolute risk aversion coefficient $\gamma_i(z)$ is an affine function in permanent income, i.e.,*

$$\gamma_i(z) \equiv \frac{u_c(c_i(z))}{u_{cc}(c_i(z))} = a_i \cdot z + b_i$$

where a_i and b_i are independent of household's permanent income z, then the first-order macro consumption Euler equation is independent of any shift in the distribution of permanent income across households.

In the homogeneous Stone-Geary utility case, the inverse of individual's absolute risk aversion is an affine function in his or her steady state consumption or permanent income, $\gamma(c_i) = \frac{1}{\sigma}(c_i - \underline{c})$. Therefore, it immediately implies that the macro IES, $\frac{1}{C^\star}\int \gamma(c_i)di = \frac{1}{\sigma}(1 - \frac{\underline{c}}{C^\star})$, is independent of the distribution of permanent income. In the heterogeneous Stone-Geary utility case, the inverse of individual's absolute risk aversion is also an affine function, $\gamma_i(c_i(z)) = \frac{1}{\sigma_i}(zC^\star - \underline{c}_i)$, where σ_i and $\underline{c}_i$ are independent of z. This implies that the macro IES is also independent of the distribution of permanent income, $\overline{\gamma} = \frac{1}{C^\star}\int \gamma_i(z)di = \mathbb{E}_I(\frac{1}{\sigma_i}) \cdot \mathbb{E}_I(z) - \frac{1}{C^\star}\mathbb{E}_I(\frac{\underline{c}_i}{\sigma_i})$.

3.5.3 The Second-order Macro Impacts: Inequality Multipliers

Although the first-order macro consumption can be neutral to permanent income inequality in the presence of non-homothetic preference over consumption, permanent income inequality may affect macro consumption to higher orders. This subsection provides a second-order characterization on macro consumption response. The second-order micro and macro consumption Euler equations are shown by the following lemma.

Lemma 2. *To the second-order, the micro and macro consumption Euler equations are,*

$$d^2\tilde{c}_{it} = \mathbb{E}_t d^2\tilde{c}_{it+1} - \frac{1}{\sigma(c_i)}\mathbb{E}_t d^2\tilde{R}_{t+1} + \frac{\sigma'(c_i)\cdot c_i}{[\sigma(c_i)]^2}\cdot \hat{c}_{it}\mathbb{E}_t\hat{R}_{t+1}$$

and

$$d^2\tilde{C}_t = \mathbb{E}_t d^2\tilde{C}_{t+1} - \left(\int \frac{c_i/C^\star}{\sigma(c_i)}di\right)\mathbb{E}_t d^2\tilde{R}_{t+1} + \left(\int \frac{\sigma'(c_i)\cdot c_i^2}{C^\star[\sigma(c_i)]^2}\hat{c}_{it}di\right)\mathbb{E}_t\hat{R}_{t+1}, \quad (3.31)$$

where the terms of $\frac{\sigma'(c_i)\cdot c_i}{[\sigma(c_i)]^2}\hat{c}_{it}\mathbb{E}_t\hat{R}_{t+1}$ *and* $\int \frac{\sigma'(c_i)\cdot c_i^2}{C^\star[\sigma(c_i)]^2}\hat{c}_{it}di\mathbb{E}_t\hat{R}_{t+1}$ *represent the second-order micro and macro "consumption-wedge shocks" respectively.*

Lemma 2 provides two properties. First, the coefficients associating the second-order terms, e.g., $d^2\tilde{c}_{it}$, $d^2\tilde{C}_t$ and others, are the same as the ones in the first-order system (see Lemma 1). Second, relative to the first-order system, the second-order system contains cross-product terms among the first-order counterparts, e.g., $\hat{c}_{it}\mathbb{E}_t\hat{R}_{t+1}$, which functions as "second-order shocks". These two properties jointly provide a huge advantage on analysis. The second-order driving process is endogenous from the perspective of the whole model, it can be treated as *exogenous* for the purpose of solving the solution to the second-order terms, since the first-order system has been solved up to this point. In fact, this advantage still preserves to any higher-order.

Imagining the first-order micro and macro systems are solved, the evolution of the second-order macro shocks $\int \frac{\sigma^{'}(c_i)\cdot c_i^2}{C^{\star}[\sigma(c_i)]^2}\hat{c}_{it}di\mathbb{E}_t\hat{R}_{t+1}$ can be characterized by the following proposition.

Proposition 4. *The second-order macro "consumption-wedge shocks" in macro consumption Euler equation,* $\int \frac{\sigma^{'}(c_i)\cdot c_i^2}{C^{\star}[\sigma(c_i)]^2}\hat{c}_{it}di\mathbb{E}_t\hat{R}_{t+1}$*, have the following relationship,*

$$\left(\int \frac{\sigma^{'}(c_i)\cdot c_i^2}{C^{\star}[\sigma(c_i)]^2}\hat{c}_{it}di\right) = \mathbb{E}_t\left(\int \frac{\sigma^{'}(c_i)\cdot c_i^2}{C^{\star}[\sigma(c_i)]^2}\hat{c}_{it+1}di\right) - \left(\int \frac{\sigma^{'}(c_i)\cdot c_i^2}{C^{\star}[\sigma(c_i)]^3}di\right)\cdot\mathbb{E}_t\hat{R}_{t+1}. \tag{3.32}$$

where $\mathcal{M}^2_{Ineq} \equiv -\int \frac{\sigma^{'}(c_i)\cdot c_i^2}{C^{\star}[\sigma(c_i)]^3}di$ *is defined as the second-order income inequality multiplier.*

Let's again use the CRRA utility and Stone-Geary utility to illustrate how the second-order income inequality multiplier works. In the CRRA utility case, since the micro IES is a constant, the first-order derivative of micro IES is zero, i.e., $\sigma^{'}(c_i) = 0$. Therefore, under CRRA specification, both of the second-order micro and macro discount-rate shocks vanish completely. This implies that permanent income inequality does not play any role to second-order or even higher orders.

In contrast, under the Stone-Geary utility specification, the second-order macro Euler equation can be represented by

$$d^2\tilde{C}_t = \mathbb{E}_t d^2\tilde{C}_{t+1} - \frac{1}{\sigma}(1 - \frac{\underline{c}}{C^\star})\mathbb{E}_t d^2\tilde{R}_{t+1} + \frac{\underline{c}}{\sigma}\left(\int \hat{c}_{it} di\right)\mathbb{E}_t\hat{R}_{t+1}$$

with

$$\left(\int \hat{c}_{it} di\right) = \mathbb{E}_t\left(\int \hat{c}_{it+1} di\right) + \mathcal{M}^2_{Ineq} \cdot \mathbb{E}_t\hat{R}_{t+1}$$

where the second-order income inequality multiplier is $\mathcal{M}^2_{Ineq} = -\frac{1}{\sigma}\left(1 - \frac{\underline{c}}{C^\star}\int z_i^{-1} di\right)$. These two equations are very useful to illustrate how non-homothetic preferences and permanent income inequality change aggregate consumption responses to the second-order. If there is a zero substance level, i.e., $\underline{c} = 0$, the second-order macro "consumption-wedge shocks" all vanish. If the distribution of permanent income becomes more disperse, then, by Jensen's inequality, $\int z_i^{-1} di$ will be larger and thus inequality multiplier $\mathcal{M}^2_{Ineq}$ will be larger as well.

To understand how aggregate consumption responds at the second-order, let's consider two types of shocks: aggregate productivity shocks on A_t and real interest rate shocks on R_{t+1}.

Aggregate Consumption Response to Productivity Shocks. Consider a positive productivity shock, $\hat{A}_t > 0$. The interest rate to the first-order rises as well as individual consumption, that's, $\hat{R}_t > 0$ and $\int \hat{c}_{it} di > 0$. This implies that the second-order consumption wedge shock is positive on impact, given the first-order responses of interest rate and the average of individual consumption. Therefore, $d^2\tilde{C}_t < 0$, which means that the second-order aggregate consumption declines. Non-homothetic preferences and permanent income inequality jointly dampens aggregate consumption responses to a productivity shock to the second-order.

Aggregate Consumption Response to Real Interest Rate Shocks. Consider a negative real interest rate shock, $\hat{R}_{t+1} < 0$. As the real interest rate declines, individual consumption rises to the first-order, that's, $\int \hat{c}_{it} di > 0$. This implies a negative correlation between the changes of interest rate and average consumption to the first-order. Therefore, a negative second-order consumption wedge shock generates an increase of aggregate consumption response to the second-order, $d^2\tilde{C}_t > 0$. As a consequence, non-homothetic preferences and permanent income inequality jointly amplifies aggregate consumption responses to a real interest rate shock. For instance at the zero-lower bound, deflation, which pushes a drop of real interest rate, will have more powerful damage in the presence of more sever income inequality.

3.6 Conclusion

This paper proposes a new method to tackle aggregation in recursive stochastic dynamic models with micro heterogeneity. This method enables us to derive sufficient statistics that can characterize explicitly how micro heterogeneity affects aggregate dynamics. To the first- and second-order, these sufficient statistics usually can be shaped by the average and the dispersion of the micro counterparts weighted by their micro impact intensities. This method can be potentially applied in different contexts to shed light on the implications of micro heterogeneity in macroeconomics.

Bibliography

Andreasen, Martin, Jesus Fernandez-Villaverde, and Juan F. Rubio-Ramirez. 2018. "The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications." *Review of Economic Studies* 85(1): 1-49.

Aiyagari, S. Rao. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving." *Quarterly Journal of Economics* 109(3): 659-84.

Artuc, Erhan, Shubham Chaudhuri, and John McLaren. 2010. "Trade Shocks and Labor Adjustment: A Structural Empirical Approach." *American Economic Review* 100(3): 1008-45.

Auclert, Adrien. 2019. "Monetary Policy and the Redistribution Channel." *American Economic Review* 109(6): 2333-67.

Baqaee, David R., and Emmanuel Farhi. 2019. "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem." *Econometrica* 87(4): 1155-203.

Barro, Robert. 1989. "The Neoclassical Approach to Fiscal Policy." in *Modern Business Cycle Theory*, R. J. Barro, ed., Cambridge, MA: Harvard University Press. pp. 178-235.

Barro, Robert, and Charles Redlick. 2011. "Macroeconomic Effects from Government Purchases and Taxes." *Quarterly Journal of Economics* 126(1): 51-102.

Baxter, Marianne, and Robert G. King. 1993. "Fiscal Policy in General Equilibrium." *American Economic Review* 83(3): 315-34.

Benigno, Gianluca, and Pierpaolo Benigno. 2003. "Price Stability in Open Economies." *Review of Economic Studies* 70(4): 743-64.

Beraja, Martin, Erik Hurst, and Juan Ospina. 2018. "The Aggregate Implications of Regional Business Cycles." Working Paper.

Bils, Mark, and Peter J. Klenow. 2004. "Some Evidence on the Importance of Sticky Prices." *Journal of Political Economy* 112(5): 947-85.

Blanchard, Olivier J., and Charles M. Kahn. 1980. "The Solution of Linear Difference Models under Rational Expectations." *Econometrica* 48: 1305-13.

Blanchard, Olivier J., and Lawrence F. Katz. 1992. "Regional Evolutions." *Brooking Papers on Economic Activity* (1): 1-61.

Borjas, George J.. 2001. "Does Immigration Grease to Wheels of the Labor Market?" *Brookings Papers on Economic Activity* (1): 69-119.

Burnside, Craig, Martin Eichenbaum, and Jonas Fisher. 2004. "Fiscal Shocks and Their Consequences." *Journal of Economic Theory* 115(1): 89-117.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. 2018. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." *Econometrica* Forthcoming.

Carroll, Christopher D. 1997. "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis." *Quarterly Journal of Economics* 112(1): 1-55.

Chari, V. V., Patrick Kehoe, and Ellen McGrattan. 2002. "Can Sticky Price Models Generate Volatile and Persistent Real Exchange Rates?" *Review of Economic Studies* 69(3): 533-63.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2012. "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review Papers & Proceedings* 101(3): 471-75.

Chodorow-Reich, Gabriel. 2018. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy* Forthcoming.

Chodorow-Reich, Gabriel, Laura Feiveson, Zachary Liscow, and William Woolston. 2012. "Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy* 4(3): 118-45.

Christinao, Lawrence, Martin Eichenbaum, and Sergio Rebelo. 2011. "When Is the Government Spending Multiplier Large?" *Journal of Political Economy* 119(1): 78-121.

Conley, Timothy, and Bill Dupor. 2013. "The American Recovery and Reinvestment Act: Solely a Government Jobs Program?" *Journal of Monetary Economics* 60(5): 535-49.

Cragg, John, and Stephen Donald. 1993. "Testing Identifiability and Specification in Instrumental Variable Models." *Econometric Theory* 9(2): 222-40.

Del Negro, Marco. 1998. "Aggregate Risk Sharing Across US States and Across European Countries." Unpublished.

Dupor, Bill, and Rodrigo Guerrero. 2017. "Local and Aggregate Fiscal Policy Multipliers." *Journal of Monetary Economics* 92: 16-30.

Dupor, Bill, and M. Saif Mehkari. 2016. "The 2009 Recovery Act: Stimulus at the Extensive and Intensive Large Margins." *European Economic Review* 85: 208-28.

Eichengreen, Barry. 1993. "European Monetary Unification." *Journal of Economic Literature* 31(3): 1321-57.

Epple, Deninis, Brett Gordon, and Holger Sieg. 2010. "A New Approach to Estimating the Production Function for Housing." *American Economic Review* 100(3): 905-24.

Farhi, Emmanuel, and Ivan Werning. 2014. "Labor Mobility within Currency Unions." Working Paper.

__ and __. 2017. "Fiscal Multipliers: Liquidity Traps and Currency Unions." In *Handbook of Macroeconomics* Volume 2, ed. by J. B. Taylor and H. Uhlig, pp. 2417-92.

Felipe, Jesus, and Franklin M. Fisher. 2003. "Aggregation in Production Functions: What Applied Economists should Know." *Metroeconomica* 54(2&3): 208-62.

Fisher, Franklin M. 1969. "The Existence of Aggregate Production Functions." *Econometrica* 37(4): 553-77.

Gali, Jordi, David Lopez-Salido, and Javier Valles, 2007. "Understanding the Effects of Government Spending on Consumption." *Journal of European Economic Association* 5(1): 227-70.

Gali, Jordi, and Tommaso Monacelli. 2008. "Optimal Monetary and Fiscal Policy in Currency Union." *Journal of International Economics* 76(1): 116-32.

Glaeser, Edward, and Joshua Gottlieb. 2009. "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States." *Journal of Economic Literature* 47(4): 983-1028.

Gonzalez, Libertad, and Francesc Ortega. 2013. "Immigration and Housing Booms: Evidence from Spain." *Journal of Regional Science* 53(1): 37-59.

Gordon, Robert, and Robert Krenn. 2010. "The End of the Great Depression: VAR Insight on the Roles of Monetary and Fiscal Policy." NBER Working Paper No.16380.

Gorman, W. M. 1953. "Community Preference Fields." *Econometrica* 21(1): 63-80.

Guren, Adam, Alisdair McKay, Emi Nakamura, and Jon Steinsson. 2018. "Housing Wealth Effects: The Long View." Working Paper.

Hall, Robert. 2009. "By How Much Dose GDP Rise if the Government Buys More Output?" *Brookings Papers on Economic Activity* Vol. 2009, No. 2, pp. 183-249.

House, Christopher, Christian Proebsting, and Linda Tesar. 2018. "The Benefits of Labor Mobility in a Currency Union." Working Paper.

Howard, Greg. 2018. "The Migration Accelerator: Labor Mobility, Housing, and Demand." Working paper.

Howard, Greg, and Carl Liebersohn. 2018. "The Geography Channel of House Price Appreciation." Working paper.

Hulten, Charles R. 1978. "Growth Accounting with Intermediate Inputs." *Review of Economic Studies* 45(3): 511-18.

Ilzetzki, E., E. Mendoza, and C. Vegh. 2013. "How big (small?) are fiscal multipliers?" *Journal of Monetary Economics* 60: 239-54.

Johnston, Michael K., Robert G. King, and Denny Lie. 2014. "Straightforward Approximate Stochastic Equilibria for Nonlinear Rational Expectations Models." CAMA Working Paper No.59/2014.

Jorda, Oscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95(1): 161-82.

Kaplan, Greg, and Sam Schulhofer-Wohl. 2017. "Understanding the Long-Run Decline in Interstate Migration." *International Economic Review* 58(1): 57-94.

Kennan, John, and James Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79(1): 211-51.

King, Robert G., and Mark Watson. 1998. "The Solution of Singular Linear Difference Systems under Rational Expectations." *International Economic Review* 39(4): 1015-26.

Leduc, Sylvain and Daniel Wilson. 2012. "Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment." In *NBER Macroeconomics Annual 2012* ed. by Acemoglu *et al*. pp. 89-142.

Leontief, W.W. 1947. "Introduction to a Theory of the Internal Structure of Functional Relationships." *Econometrica* 15(4): 361-73.

Lombardo, Giovanni, and Alan Sutherland. 2007. "Computing Second-order-accurate Solutions for Rational Expectation Models: Using Linear Solution Methods." *Journal of Economic Dynamics and Control* 31: 515-30.

Lucas, Robert, and Edward Prescott. 1974. "Equilibrium Search and Unemployment." *Journal of Economic Theory* 7(2): 188-209.

Molloy, Raven, Christopher Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspective* 25(3): 173-96.

Moretti, Enrico. 2011. "Local Labor Markets." *Handbook of Labor Economics* Vol. 4, pp. 1237-313.

Mundell, Robert. 1961. "A Theory of Optimum Currency Areas." *American Economic Review* 51(4): 657-65.

Nataf, A. 1948. "Sur la possibilité de construction de certains macromodèles." *Econometrica* 16 (3): 232-44.

Nakamura, Emi, and Jon Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104(3): 753-92.

Ottaviano, Gianmacro, and Giovanni Peri. 2006. "Wages, Rents and Prices: The Effects of Immigration on US Natives." Working Paper.

Owyang, Michael, Valerie Ramey, and Sarah Zubairy. 2013. "Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data." *American Economic Review Papers & Proceedings* 103(3): 129-34.

Piazzesi, Monika, and Martin Schneider. 2016. "Housing and Macroeconomics." In *Handbook of Macroeconomics* Volume 2, ed. by J. B. Taylor and H. Uhlig, pp. 1547-640.

Prescott, Edward C., and Rajnish Mehra. 1980. "Recursive Competitive Equilibrium: The Case of Homogeneous Households." *Econometrica* 48(6): 1365-79.

Ramey, Valerie. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics* Volume 2, ed. by J. B. Taylor and H. Uhlig, pp. 71-162.

Redding, Stephen, and Esteban Rossi-Hansberg. 2017. "Quantitative Spatial Economics." *Annual Review of Economics* 9: 21-58.

Robinson, Joan. 1953. "The Production Function and the Theory of Capital." *Review of Economic Studies* 21(2): 81-106.

Saiz, Albert. 2007. "Immigration and Housing Rents in American Cities." *Journal of Urban Economics* 61(2): 345-71.

Shoag, Daniel. 2015. "The Impact of Government Spending Shocks: Evidence on the Multiplier from State Pension Plan Returns." Working Paper.

Sims, Christopher A. 2002. "Solving Linear Rational Expectations Models." *Computational Economics* 20(1-2): 1-20.

Solow, Robert. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics* 39: 312-20.

Stock, James, and Mark Watson. 2018. "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments." NBER Working Papers No. 24216.

Stock, James, and M. Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg.* edited by D. Andrews and J. Stock, 80-108. Cambridge. UK: Cambridge University Press.

Straub, Ludwig. 2018. "Consumption, Savings, and the Distribution of Permanent Income." Working Paper.

Suares Serrato, Juan Carlos and Philippe Wingender. 2016. "Estimating Local Fiscal Multipliers." Working Paper.

Wicksell, Knut. 1893. "Über Wert, Kapital und Rente nach den neueren nationalökonomischen Theorien."

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

Curriculum Vitae

[redacted]