Regional Heterogeneity and Aggregate Fiscal Multiplier

Man Chon (Tommy) Iao*

October 30, 2024

Abstract

I examine the role of household heterogeneity in the transmission of government expenditure. Exploiting regional variation in military procurement spending in the US, I find that the local multipliers of government spending are negatively correlated with the share of hand-to-mouth in the region. Econometrically, heterogeneous effects across regions introduce a bias to previous estimates of the average local multiplier. Correcting for the bias reduces the average local multiplier to below 1 and brings back the absent inflationary response. Calibrated to the average share of hand-to-mouth, a monetary union TANK model can reproduce the unbiased estimates. Lastly, using contract-level data, I present evidences on the transmission mechanisms of military spending and argue that the negative relationship between the local multipliers and the share of hand-to-mouth is driven by the special composition of military spending.

*Contact: mci246@nyu.edu

1 Introduction

Cross-regional identification has become a popular approach in the empirical study of fiscal multipliers. A main advantage of this approach is the wealth of plausibly exogenous variation in fiscal policy across regions, allowing for a clean reduced-form identification of fiscal multipliers.[1] On the other hand, a burgeoning literature on fiscal multipliers in heterogeneous agent models has suggested that household heterogeneity plays a key role in the transmission of fiscal policy and hence affects the sizes of fiscal multipliers.[2] Given the well-documented regional heterogeneity in demographics in the US, the theory suggests that the effect of government spending might vary substantially across states. In this paper, I ask

1. Does the effect of government spending vary across states in the US?
2. How does regional heterogeneity affect the cross-regional identification method?
3. What's the implication of the *cross-regional relationship between fiscal multipliers and household heterogeneity*?

In light of the HANK literature, I focus on a particular type of regional household heterogeneity, namely the fraction of hand-to-mouth (hereafter HtM) households (Kaplan, Violante and Weidner 2014). Combining the wealth data from the Survey of Consumer Finance and the demographic data from decennial Census and the American Community Survey (ACS), I uncover a substantial regional variation in the fraction of HtM in the U.S.

I then explore the relationship between the fraction of HtM and the effectiveness of government spending using cross-regional identification method. I exploit variation in state military procurement spending using a shift-share design as in Nakamura and Steinsson (2014) and allow the fiscal multiplier to vary with the fraction of HtM. I find that controlling for heterogeneous effect decreases the "open-economy relative multipliers" by about 30%, compared to the estimates reported in Nakamura and Steinsson (2014). This result is robust across a wide range of specifications. I provide a simple econometric theory that reconciles the two estimates – in short, when heterogeneous effect presents and is not controlled, shift-share design requires an extra assumption for identifying the *average* multiplier.[3]

[1] For instance, Nakamura and Steinsson (2014) uses a shift-share design to identify exogenous variations in states' military spending. Their identifying assumption is that the national military spending does not increase when the economies of some states are doing poorly relative to other states–a very plausible assumption. See Chodorow-Reich (2019) for a survey of cross-regional identification schemes.

[2] See for example, Galí, López-Salido and Vallés (2007), Brinca et al. (2016), Auclert, Rognlie and Straub (2018), Bilbiie (2020), Broer, Krusell and Öberg (2021), Cantore and Freund (2021), Basso and Rachedi (2021)

[3] The bias caused by heterogeneous treatment effect has also been discovered in other empirical studies. For example, Orchard et al. (2023) find that heterogeneity biases upward the conventional TWFE estimate of average MPC by 19% in the famous study by Parker et al. (2013)

The econometric bias has important macro implications. I develop a monetary union TANK model (Farhi and Werning 2016) that is calibrated to the average share of HtM and show that it can reproduce the unbiased average local multiplier if the dividend is distributed uniformly. Furthermore, since the average local multiplier is now smaller, the model does not require GHH preference as in Nakamura and Steinsson (2014). Thus, correcting the bias reconciles the empirical evidence and Auclert et al. (2023)'s result that GHH preference implies implausibly high fiscal mulitpliers.

Interestingly, the empirical results also suggest that government spending is *less* stimulative in states with *more* wealthy HtM. This correlation cannot be explained by other regional heterogeneity such as income, homeownership, and value-added share of manufacturing. To understand the mechanism underlying the correlation, I examine regional pattern of the local responses to aggregate shocks. I find that states with more wealthy HtM are indeed *more* procyclical, suggesting that the correlation is not driven by common mechanisms such as regional heterogeneity in aggregate labor supply and that the transmission mechanism of military procurement spending may be special.

Utilizing contract-level data, I document preliminary evidences on the transmission of military procurement spending. First, the composition of the spending is highly skewed – three 3-digit NAICS industries (Manufacturing of transportation equipment (336), Professional, Scientific, and Technical Services (541), and Manufacturing of computer and electronic products (334)) together account for more than 60% of the spending. For comparison, their value-added share of GDP is only 10% in total. Second, military procurement spending has larger direct effect on these three industries but spending on these industries have lower aggregate effects on the local economy. Overall, the evidences support the notion that the transmission mechanism of military procurement spending is unique. As a result, the fiscal multpliers identifed using variation in military spending are not straightforward to carry over to other types of governement spending.

Related literature This paper contributes to the literature on the heterogeneous effects of government spending. The most related paper is Basso and Rachedi (2021), who studied how the age structure of the economy affects the effectiveness of government spending using a similar empirical design. Other related papers include Juarros (2021), who focused on the employment share of small businesses, and Demyanyk, Loutskina and Murphy (2019), who looked at consumer indebtness. Instead, this paper focuses on the share of HtM which is directly related to the MPC mechanism embedded in HANK models, thereby providing an empirical test of the MPC mechanism.

The rest of the paper is organized as follows. Section 2 describes the construction of the state-level HtM measures. Section 3 discusses the potential bias introduced by heterogeneous effect and presents the main estimation results regarding the average multiplier and the correlation between the HtM measure and the multiplier. Section 4 examines the heterogeneous local responses to aggregate shocks. Section 5 presents preliminary evidence on the transmission of military spending. Section 6 concludes.

2 Regional heterogeneity in the share of HtM

Identifying HtM households requires detailed information on households' balance sheets. In the US, the only publicly available dataset that satisfies this requirement and also fits into the sample period is the Survey of Consumer Finance.[4] Unfortunately, the SCF does not disclose the residential state of the sample households so I cannot directly compute the fraction of HtM households for each state. Instead, I use an imputation approach that combines the SCF with the state-level census data. The approach involves three steps:

1. Identify HtM households in the SCF.
2. Estimate a prediction model for HtM status based on an extensive set of demographic and income variables.
3. Run the prediction model on the census data to obtain a state-level HtM measure.

In the followings, I explain each step in details and discuss the imputation results.

2.1 Identifying HtM households in SCF

In general, being HtM means holding zero wealth at the end of a payment period. Since in the data we only observe the household's balance sheet at a random point in time or just the average balance over a period, there are multiple reasonable definitions of HtM.[5] For simplicity and transparency, I use the baseline definition in Kaplan, Violante and Weidner (2014) which defines a household as HtM if either one of the following conditions holds:

1. The liquid wealth balance is positive but are equal to or less than half of their biweekly earnings.

[4] The Panel Study of Income Dynamics (PSID) also satisfies this requirement after the redesign in 1999 and indeed Aguiar et al. (2021) studies the consumption behavior of HtM households using PSID. However, the short sample period and the small sample size do not fit into the imputation approach here so PSID is not considered.

[5] See Kaplan et al. (2014) for an extensive discussion of the different definitions and their implications.

2. The liquid wealth balance is negative but are equal to or less than half of their biweekly earnings less the credit limit which is equal to the monthly earnings by assumption.

I further distinguish the HtM households by their illiquid wealth balance. A HtM household is called *wealthy* HtM if its illiquid wealth balance is positive and is called *poor* HtM otherwise. See Appendix B for the exact definition of wealth and income.

Following Kaplan, Violante and Weidner (2014), I exclude households with negative income or only self-employment income. I further restrict the sample to households in which the head is between 22 and 64 years old. Figure 1 plots the share of each type of HtM households in the SCF for each sample year. Over the years, about 20% of households are wealthy HtM, while 13% of households are poor HtM. The fractions are quite stable over time until 2016 when we see a large drop in the fraction of poor HtM households. Whether this change is transitory or structural merits further exploration.

Table 1 reports some summary statistics for each group of households. The top panel reports the average age and the fraction of households in each selected demographic groups, where the demographic refers to the head of the household. We can see that the poor HtM households are starkly different from the other two groups: they are younger, less likely to be white, have a bachelors degree, or own a house.[6] On the other hand, wealthy HtM households are quite similar to non-HtM households in their demographic composition, except in their education levels. The bottom panel reports median income and wealth balances for each group. Not surprisingly, poor HtM households have the lowest income and the highest mortgage-payment-to-income ratio, while wealthy HtM households are in the middle of the three groups. Overall, the summary statistics suggest that it is crucial to incorporate altogether demographic, income, and mortgage payment information into a prediction model.

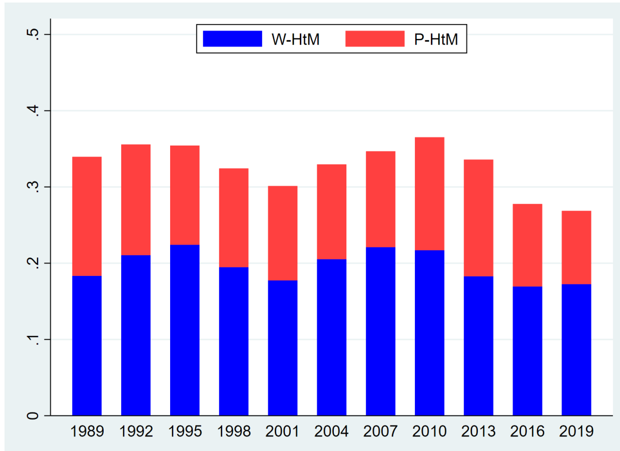
2.2 Estimating a prediction model for HtM status

I use multinomial logistic regression as my prediction model. The set of predictor variables consist of a third-order polynomial of log income and log mortgage payment as well as an extensive set of categorical variables including sex, age, marital status, race, homeownership, mortgage status, education, and whether the household has investment income. Precisely, I estimate the following parametric model with MLE:

$$\log \frac{\mathbb{P}(\text{HtM}_i = h \mid X_i)}{\mathbb{P}(\text{HtM}_i = \text{N-HtM}) \mid X_i)} = \sum_{k=1}^{3} \beta^h_{y,k} (\log y_i)^k + \sum_{k=1}^{3} \beta^h_{mp,k} (\log mp_i)^k + Z_i' \gamma^h$$

[6]Homeowners can be poor HtM if they have negative home equity.

Figure 1: Fraction of HtM households by year



NOTE. The blue bar corresponds to the fraction of wealthy HtM and the red bar corresponds to the fraction of poor HtM.

where $h \in \{\text{W-HtM, P-HtM}\}$ is the HtM status, y_i is labor income, mp_i is mortgage payment, and Z_i is a vector of demographic variables. All dollar-valued variables are deflated to 1999 dollars using CPI-U. When estimating the model, I drop households whose labor income or mortgage payment is at the top 1% of the distribution and restrict the sample to years before 2010 to match the sample period in the main regression analyses.

Figure 2 plots the predicted fraction of HtM households against the true value for each year. The prediction model performs well for the sample period 1989 - 2010 but over predicts for year 2013 onward. See Appendix C for a bootstrap analysis of the accuracy of the prediction model.

2.3 Predicting fraction of HtM households for each state

With the estimated prediction model in hands, we can feed in state census data to obtain an estimate of the fraction of HtM households in each state. The sample period of the main analysis is 1966-2006 so naturally I should use the 1960-2010 census data. However, the mortgage payment variable is only available since 1980 so I decide to use the 1980-2010 data instead. More specifically, I use the 1980, 1990, and 2000 decennial census and 2001-2010 American Community Survey. For each household in the sample I compute the predicted probabilities of each H2M status and for each state-year I aggregate the probabilities to obtain an estimate of the fraction of HtM households.

The state-level HtM measures are constructed by taking (weighted) average of the predicted fractions of HtM households over the years. I put 1/10 weight on the ACS samples

Table 1: Summary statistics by HtM status

Mean:	Age	Homeowner	White non-hispanic	Bachelor degree
Wealthy HtM	43.5	0.74	0.67	0.22
Poor HtM	39.0	0.05	0.47	0.07
Not HtM	43.6	0.69	0.74	0.39
Median:	**Labor earnings**	**Mortgage PTI**	**Liquild wealth**	**Illiquid wealth**
Wealthy HtM	34273	0.19	0	32563
Poor HtM	14934	0.21	0	0
Not HtM	44954	0.16	5558	64506

NOTE. The top panel reports (mean) demographic statistics for each group of households. The first column is the average age and each of the remaining columns reports the shares of households in a particular demographic group, conditional on HtM status. The bottom panel reports (median) income and wealth statistics as well as mortgage-payment-to-income ratio (second column). All dollar-valued variables are deflated to 1999 dollars using CPI-U.

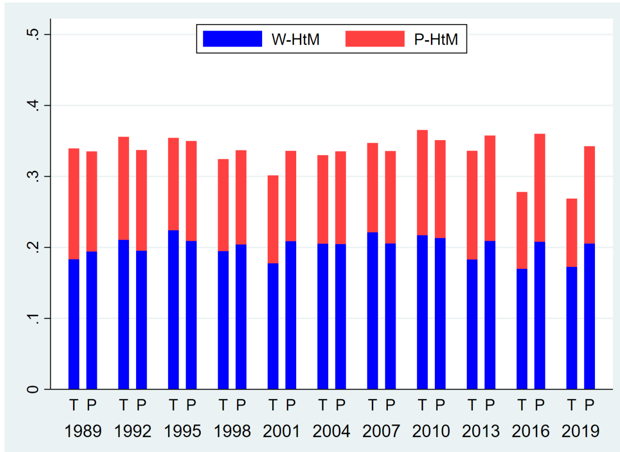
to adjust for over-sampling in the 2000s. Table 2 reports the summary statistics of the state-level HtM measures. The (unweighted) mean of the HtM measures is quite close to the national average computed from SCF. Also, there are substantial heterogeneity across states, with the lowest state (Alaska) having 26% of households being HtM and the highest state (Mississippi) having 42%. Interestingly, there is no correlation between the wealthy and the poor HtM measure. This implies that the regional variation in the fraction of wealthy HtM is at the HtM vs. non-HtM margin. Figure 3 displays a heatmap for the state-level HtM measures. Geographically, states with higher fraction of wealthy HtM households are typically in the central and south regions, while the fractions of poor HtM households are more uniformly distributed across states. Note that some states (e.g. California, Hawaii) have the lowest fraction of wealthy HtM but highest fraction of poor HtM, reassuring the importance of distinguishing the two.

Table 2: Summary statistics of the state-level HtM measures

VARIABLES	N	mean	sd	min	max	Corr(·, W-HtM)	Corr(·, P-HtM)	Corr(·, HtM)
W-HtM	51	0.2217	0.0230	0.1708	0.2775	1		
P-HtM	51	0.1263	0.0215	0.0865	0.2084	0.0167	1	
HtM	51	0.3480	0.0317	0.2591	0.4235	0.7358	0.6893	1

NOTE. The statistics are computed without using state population as weights.

Figure 2: Predicted fraction of HtM households in SCF by year



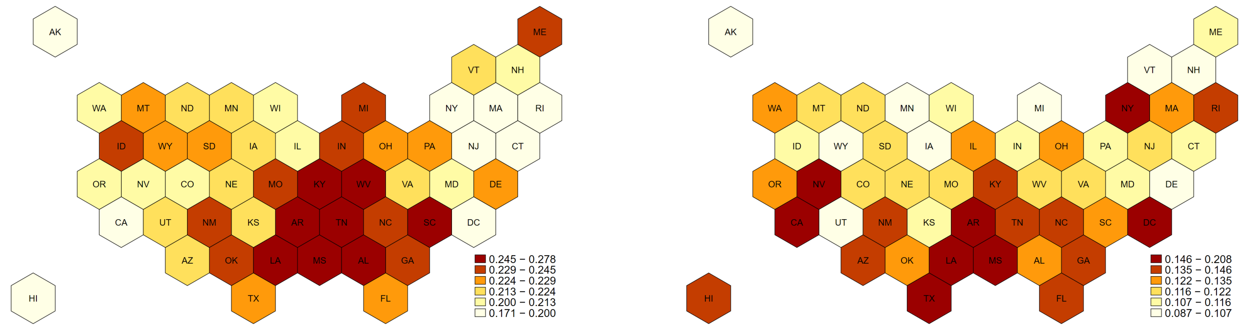
NOTE. The "T" bar is the true value and the "P" bar is the predicted value. The year 2013 - 2019 are out-of-sample prediction.

2.4 What explains state-level variation in the share of HtM?

Since the state-level HtM measures are constructed using the demographic information of households in each state, they are necessarily correlated with state-level demographic statistics, causing a concern of spurious relationship. In Table 3, I report results from regressing the wealthy HtM measures over a set of demographic statistics. All the statistics are computed using the same sample as in the imputation of the HtM measures. Not surprisingly, income and wealth (proxied by homeownership) are strong predictors for the wealthy HtM measure. Together they can explain 93% of the state-level variation in the share of wealthy HtM and indeed the average labor earnings itself can already explain 70% of the variation. To alleviate concern of spurious correlation, in my empirical analyses I perform extensive robustness tests to show that the economic relevance of the HtM measures is not driven by income, homeownership, nor any other unidimensional demographic factor.

3 Local multiplier and the share of HtM

Given the substantial regional variation in the share of HtM, theory suggests that the effect of governement spending may vary across regions. In this section, I study the empirical relationship between the share of HtM and the effect of governement spending and discuss the consequences of heterogeneous effects to the cross-regional identification method.



(a) W-HtM **(b)** P-HtM

Figure 3: Predicted fraction of HtM households by states

Table 3: Demographics vs. HtM

	(1) WHtM	(2) WHtM	(3) WHtM	(4) WHtM	(5) WHtM
PHtM	0.02	-0.68***	0.55***	0.55***	0.60***
	(0.20)	(0.13)	(0.12)	(0.10)	(0.10)
log(labor earnings)		-0.15***	-0.06***	-0.05***	-0.05***
		(0.01)	(0.01)	(0.01)	(0.01)
homeownership			0.37***	0.35***	0.36***
			(0.03)	(0.03)	(0.03)
College				-0.04	-0.04
				(0.03)	(0.03)
Young ($22 \leq$age≤ 30)					0.07*
					(0.04)
Constant	0.22***	1.79***	0.53***	0.48***	0.41***
	(0.02)	(0.14)	(0.14)	(0.14)	(0.13)
Observations	51	51	51	51	51
R-squared	0.00	0.70	0.93	0.94	0.94

3.1 Empirical design and identification

Identification of fiscal multiplier requires exogenous variation in government spending. Following Nakamura and Steinsson (2014), I focus on military procurement spending across states in the U.S. and use Bartik IV to identify exogenous variation. This strategy has been widely adopted in the literature, but only a few of the previous researches consider the possibility of heterogeneous multipliers, which is the heart of my empirical exercise.[7]

Before discussing the main empirical specification, let's start with the general econometric

[7] The only two exceptions I know of are Basso and Rachedi (2021) and Juarros (2021). The former studies the relationship between the share of young workers and the effect of governement spending, while the latter focuses on the employment share of small firms.

model of regional regression with heterogeneous effects:

$$Y_{it} = \beta_i G_{it} + \alpha_i + \alpha_t + \varepsilon_{it} \tag{1}$$

$$G_{it} = \gamma_i G_t + \alpha_i^G + \alpha_t^G + \nu_{it} \tag{2}$$

where Y_{it} is some outcome variable of interest (e.g., output growth), G_{it} is regional governement spending, G_t is national governement spending, and the α's are fixed effects. Note that the coefficient on G_{it}, β_i, is region-specific and represents the causal effect of governement spending in the region. If instead β_i is common across regions, then this specification reduces to the usual Bartik-TWFE setup. Equation (2) is the first-stage of the Bartik IV. The coefficient γ_i is the sensitivity of region i's governement spending to national governement spending. The standard identifying assumption is $\mathbb{E}[\gamma_i G_t \varepsilon_{it} \mid \alpha_i, \alpha_t, \alpha_i^G, \alpha_t^G] = 0$, meaning that the national government spending does not increase when the economies of regions with high sensitivities are doing poorly relative to other regions.

We are interested in the estimand of the 2SLS estimator in this specification when a researcher agnostically assumes homogeneous effect $\beta_i = \beta$. In Appendix D.1, I show that the estimand β^{TWFE} can be decomposed into three components:

$$\beta^{TWFE} = \underbrace{\mathbb{E}[\beta_i]}_{ATE} + \underbrace{\mathbb{E}\left[\frac{\tilde{\gamma}_i^{\,2}}{\mathbb{E}[\tilde{\gamma}_i^{\,2}]} \cdot \tilde{\beta}_i\right]}_{Bartik} + \underbrace{\frac{\mathrm{Cov}(\beta_i, \gamma_i)}{\mathbb{E}[\tilde{\gamma}_i^{\,2}]} \cdot \left(\mathbb{E}[\gamma_i] + \frac{\mathbb{E}[\hat{G}_t \hat{\alpha}_t^G]}{\mathbb{E}[\hat{G}_t^2]}\right)}_{OVB}$$

where $\tilde{x}$ denotes within-time deviation and $\hat{x}$ denotes within-unit deviation. The first term is the average multiplier. The second term is the OLS adjustment caused by the use of Bartik IV. It is a convex combination of the (deviated) multiplier with weights given by the square of the sensitivity.[8] Intuitively, the square of the sensitivity measures the fraction of variation in the identifed shock that a region has contributed to. To maximally explain the variation in the outcome variable, OLS puts more weight on regions that contribute more to the variation in the identifed shock. Finally, the last term is the omitted variable bias due to the inclusion of time fixed effect. To see where it comes from, substitute equation (2) into equation (1) to obtain the reduced-form equation:

$$Y_{it} = \beta_i \gamma_i G_t + (\alpha_i + \beta_i \alpha_i^G) + (\alpha_t + \beta_i \alpha_t^G) + (\varepsilon_{it} + \beta_i \nu_{it})$$

[8] Goldsmith-Pinkham, Sorkin and Swift (2020) have derived a similar expression in their section 4. Their expression did not involve the OVB term because they had assumed that $\mathrm{Cov}(\beta_i, \gamma_i) = 0$ (Assumption 3.2). As I will show in the next section, this assumption is violated in my empirical exercise so it is important to characterize it. See also Sun and Shapiro (2022) for a discussion of potential bias in shift-share design caused by heterogeneous treatment effects.

Note that there is a interactive fixed effect term $\beta_i \alpha_t^G$. Intuitively, when there is an overall boom in governement spending, regions with different multipliers are affected differently. The time fixed effect cannot fully absorb this heterogeneous response to aggregate shock, therefore leaving an omitted variable.

Suppose we are interested in estimating the average multiplier.[9] Then the bias of the 2SLS estimator can be attributed to the Bartik term and the OVB term. These two terms are characteristically different. The size of the Bartik term mainly depends on the skewness of the Bartik shift-share distribution, while the size of the OVB term is determined by the correlation between the Bartik shift-share and the multiplier. This decomposition will help us understand why my empirical results differ from previous results in the literature. I defer the discussion to Section 3.4.

3.2 Main specification and estimation results

In order to control for the potential bias due to heterogeneous multpliers and to estimate the relationship between the multiplier and the share of HtM, I consider the following econometric specification:

$$\begin{aligned}\frac{Y_{it} - Y_{it-2}}{Y_{it-2}} = \beta \frac{G_{it} - G_{it-2}}{Y_{it-2}} &+ \delta_w \log(WHtM_i) \times \frac{G_{it} - G_{it-2}}{Y_{it-2}} \\ &+ \delta_p \log(PHtM_i) \times \frac{G_{it} - G_{it-2}}{Y_{it-2}} + \alpha_i + \alpha_t + \varepsilon_{it}\end{aligned}$$

where Y_{it} is per capita output in state i in year t, G_{it} is per capita military procurement spending in state i in year t, and $\log(WHtM_i)$ and $\log(PHtM_i)$ are log of the constructed HtM measures. The log-HtM-measures are de-meaned and multiplied by 100 before estimation for ease of interpretation. If the multpliers are heterogeneous and the HtM measures can fully explain the heterogeneity, then the coefficients δ_w, δ_p on the interaction terms are non-zero and β is equal to the average multiplier. I estimate the equation with 2SLS, instrumenting both the government spending term and the interaction terms with Bartik-IV as discussed in the last section. Specifically, the first statge is

$$\frac{G_{it} - G_{it-2}}{Y_{it-2}} = \gamma_i \frac{G_t - G_{t-2}}{Y_{t-2}} + \alpha_i^G + \alpha_t^G + \varepsilon_{it}^G$$

[9] The current practice in the literature is to pair the estimted local multiplier with a two-region monetary-union model to infer the aggregate multplier. The estimated multiplier is treated as an "identifed moment" that helps pin down the correct model (Nakamura and Steinsson 2018). If we calibrate the home region in the model to resemble the average region in the data, then necessarily the model implied local multiplier corresponds to the average multiplier in the regression model and hence this is the moment of interest.

Table 4: The Heterogeneous Effects of Military Spending

	Output		Output defl. state CPI		Employment		CPI	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\frac{G_{it}-G_{it-2}}{Y_{it-2}}$	1.43***	1.07***	1.34***	0.94***	1.28***	1.08***	0.03	0.14
	(0.36)	(0.37)	(0.36)	(0.32)	(0.29)	(0.25)	(0.18)	(0.13)
$\log(WHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$		-0.08*		-0.09***		-0.05***		0.04***
		(0.04)		(0.03)		(0.02)		(0.01)
$\log(PHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$		-0.01		-0.01		0.00		0.01
		(0.02)		(0.02)		(0.01)		(0.01)
Observations	1,989	1,989	1,989	1,989	1,989	1,989	1,763	1,763

NOTE. The odd columns present the original results in Table 2 of Nakamura and Steinsson (2014). The government spending variable is per capita military procurement spending. The HtM measures are all standardized. Standard errors are clustered by state. *** p<0.01, ** p<0.05, * p<0.1

where Y_t is national per capita output in year t and G_{it} is national per capita military procurement spending in year t. As in Nakamura and Steinsson (2014), the set of instruments is $\{D_i \frac{G_t - G_{t-2}}{Y_{t-2}}\}_i$. Note that it can be used to instrument also the interaction terms because the HtM measures are time-invariant. For robustness, I consider other econometric specification such as interactive fixed effect in Section 3.3. I also estimate the effects on employment and inflation by replacing the dependent variable with 2-year growth rate of employment rate and 2-year inflation rate. The panel spans the period 1966-2006 and consists of the 50 states and D.C. The dataset is from the replication material of Nakamura and Steinsson (2014) and I refer the reader to the paper for the construction of the dataset.

Table 4 reports the regression results. For comparison, I also report the original results of Nakamura and Steinsson (2014) in the odd columns. After accounting for heterogeneous multipliers, both the output multipliers and the employment multiplier drop substantially while the inflation multiplier increases by an order of magnitude. In particular, the state-CPI-adjusted output multiplier (column 4) is now 0.94, about 30% smaller than the original estimate of 1.34. This means on average one-dollar increase in military procurement spending causes only 94-cent increase in output, suggesting a crowd-out effect of governement spending which is typically found from the VAR analyses of the effect of national governement spending (Ramey 2011).

The coefficients on the interaction term with poor HtM measure are always close to 0 and statistically insignificant, while the coefficients on the interaction term with wealthy HtM measure are large and significant in all specifications. Thus, we can conclude that it is the heterogeneity in the fraction of wealthy HtM that explains the heterogeneous multipliers. Surprisingly, theses coefficients are negative in the output and employment regressions but positive in the CPI regression, meaning that states with proportionally more wealthy HtM households have lower fiscal multipliers but larger inflation responses. For the state-CPI-adjusted output multplier (column 4), a 1% increase in the wealthy HtM measure relative to the cross-sectional mean is associated with 0.09 decrease in the local multplier. The implied inter-quartile range is [0.3, 1.47].

3.3 Robustness

The correlation of the HtM measures with other economic factors that affect the transmission of governement spending can bias the estimates. To address this concern, I run "horse-race" regressions where once at a time I add an extra interaction term to control for the confounder. For simplicity, I focus on the state-CPI-adjusted output multiplier (column 4 of Table 4) and keep only the wealthy HtM measure because the result is most salient in this specification. I consider both demographic factors and supply-side factors as confounding factors.

Household heterogeneity The selected demographic factors are 1) share of 20-29 years old white male, 2) share of college graduates, 3) share of homeowners, 4) share of households with mortgage, and 5) mean household labor income. These factors are selected either because it is documented to be correlated with the local multiplier or it is strongly correlated with the wealthy HtM measure. All factors are the time average calculated using the 1980-2000 census and 2001-2010 ACS (same as the HtM measures), except the share of 20-29 years old white male which is independently constructed by Basso and Rachedi (2021). Table 5 presents the results. Column 1 is a re-estimation of the main specification, keeping only the wealthy HtM measure. Not surprisingly, the result barely changes as the poor HtM measure was estimated to be unimportant for the local multplier. Column 2 adds the share of 20-29 years old white male which Basso and Rachedi (2021) have found to be positively correlated with the local multiplier using a similar empirical design. My result corroborates their finding, and most importantly, the correlation between the share of wealthy HtM and the local multplier persists. The average multplier slightly increases but is still smaller than the original estimate.

Column 3-6 present results from adding one-by-one a demographic variable that best explains the variation of the wealthy HtM measure. Except for the share of college graduates (column 3), adding an extra demographic variable has marginal effect on the main result and the demographic variable does not help to explain the local multplier. On the other hand, the share of college graduates is estimated to be negatively correlated with the local multiplier and adding it change the result significantly. The magnitude of the coefficient on the wealthy HtM measure is larger and the average multiplier increases. This result suggests that the share of college graduates is also correlated with the local multplier and

Table 5: Robustness: Household Heterogeneity

LHS = output defl. state CPI	(1) Baseline	(2) Basso-Rachedi	(3) College	(4) Homeowner	(5) Mortgage	(6) Income
$\frac{G_{it}-G_{it-2}}{Y_{it-2}}$	0.96***	1.03***	1.21***	0.98***	1.05***	1.00***
	(0.306)	(0.299)	(0.341)	(0.312)	(0.298)	(0.290)
$\log(WHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$	-0.09***	-0.09***	-0.17***	-0.09**	-0.09***	-0.11*
	(0.034)	(0.031)	(0.058)	(0.039)	(0.031)	(0.061)
$\text{Factor}_i \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$		0.02	-0.05*	-0.00	-0.03	-0.01
		(0.030)	(0.029)	(0.021)	(0.025)	(0.034)
Observations	1,938	1,938	1,938	1,938	1,938	1,938

Note. The LHS variable is state-CPI-adjusted output (column 3-4 of Table 4). Column (1) is the baseline result where only the wealthy HtM measure is included. Column (2) adds the share of 20-29 years old white male reported in Basso and Rachedi (2021). Column (3) adds the share of college graduates. Column (4) adds the share of homeowners. Column (5) adds the share of households with mortgage. Column (6) adds the mean household labor income. All the demographic variables are log-transformed, de-meaned, and multiplied by 100. Standard errors are clustered by state. *** p<0.01, ** p<0.05, * p<0.1

that the previous estimate has picked up some of the correlation because college education is negatively correlated with the wealthy HtM measure.

Sectoral composition and trade openness The wealthy HtM measure might also be correlated with the sectoral composition of value added in each state. This can bias my estimate as the literature have documented that the effect of governement spending varies across sectors (Bouakez et al. 2022). I consider three sectors, namely manufacturing, construction, and service, because they are economically large and are significantly affected by military procurement spending (Nakamura and Steinsson 2014). On the other hand, since the degree of trade openness is a determinant of the local multplier (Farhi and Werning 2016), its correlation with the wealthy HtM measure can also bias the estimate. I compute the fraction of across-state shipment (in values) using the 2002 US Commodity Flow Survey and use it as a proxy for trade openness. To sum up, the set of confounding factors includes 1) value-added share of manufacturing, 2) value-added share of construction, 3) value-added share of service, and 4) the fraction of across-state shipment.

Table 6 presents the results. Overall, controlling for supply-side factor does not change the result. The only factor that helps explain the local multiplier is the value-added share of service. Controlling for it strengthens the correlation between the wealthy HtM measure and the local multplier. The average multiplier slightly increases to around 1 in all specifications, still substantially lower than the original estimate of 1.34.

Table 6: Robustness: Sectoral Composition and Trade Openness

LHS = output defl. state CPI	(1) Baseline	(2) Openness	(3) Manufacturing	(4) Construction	(5) Service
$\frac{G_{it}-G_{it-2}}{Y_{it-2}}$	0.96***	0.98***	1.02***	1.01***	1.01***
	(0.306)	(0.306)	(0.325)	(0.296)	(0.304)
$\log(WHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$	-0.09***	-0.09***	-0.09***	-0.10***	-0.12***
	(0.034)	(0.033)	(0.029)	(0.036)	(0.031)
$\text{Factor}_i \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$		-0.00	-0.00	0.01	-0.06**
		(0.013)	(0.010)	(0.022)	(0.027)
Observations	1,938	1,938	1,938	1,938	1,938

NOTE. The LHS variable is state-CPI-adjusted output (column 3-4 of Table 4). Column (1) is the baseline result where only the wealthy HtM measure is included. Column (2) addes the share of between-state shipments calculated using the 2002 US Commodity Flow Survey. Column (3)-(5) add the output share of each selected sector.

Other robustness checks As shown in Section 3.1, the skewness of the Bartik shift-shares will affect the size of the bias if the multplier heterogeneity is not controlled. In other words, if the HtM measure has captured most of the variation in the multipliers, then using a different set of Bartik shift-shares should not change the result. Following Nakamura and Steinsson (2014), I compute the average fraction of military spending in each state during 1966-1971 as another set of Bartik shift-shares and re-estimate the output and employment multiplier (column 2 and 6 of Table 4).[10] Another concern is the heterogeneity in each state's response to aggregate shocks. For instance, if states with more HtM respond more strongly to oil price shock and national military spending is correlated with the shock, then my estimate of the correlation between the HtM measure and the local multplier is upward biased. To address this concern, I add an interaction term of oil prices or real interest rates with state fixed effect to the output multiplier regression. Table 7 reports the results. Column 1-4 are the results from using the new set of Bartik shift-shares. Without controlling for heterogeneous multipliers, the estimates of the average multiplier inflate significantly because this new set of Bartik shift-shares is more skewed. In stark contrast, when the HtM measures are included, the average multpliers are remarkedly stable across the two sets of Bartik IV. This is an evidence that the Bartik bias is present and the HtM measures capture well the multiplier heterogeneity. On the other hand, allowing each state to respond differently to aggregate oil prices and real interest rates only changes the results slightly. This is because national military spending is only weakly correlated with oil prices and real interest rates. Indeed,

[10]To construct an instrument for the interaction term, I interact the Bartik-IV with the HtM measure.

Table 7: Robustness: Alternative Specifications

	Output level instr.		Employment level instr.		Output with oil controls		Output with real int. controls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\frac{G_{it}-G_{it-2}}{Y_{it-2}}$	2.48***	1.00	1.81***	0.86**	1.32***	0.92**	1.40***	1.06***
	(0.94)	(0.71)	(0.41)	(0.36)	(0.36)	(0.41)	(0.35)	(0.36)
$\log(WHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$		-0.10*		-0.06***		-0.08*		-0.07*
		(0.05)		(0.02)		(0.05)		(0.04)
$\log(PHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$		-0.02		0.01		-0.01		-0.01
		(0.02)		(0.01)		(0.02)		(0.02)
Observations	1,989	1,989	1,989	1,989	1,989	1,989	1,989	1,989

NOTE. The odd columns present the original results in Table 3 of Nakamura and Steinsson (2014). Column 1-4 use national military spending scaled by fraction of military spending in the state in 1966–1971 relative to the average fraction as the instrument. Column 5-6 add price of oil interacted with state dummies as controls. Column 7-8 add real interest rate interacted with state dummies as controls. Standard errors are clustered by state. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

states with more wealthy HtM are more procyclical in general, a point that we will further discuss in Section 4.

Finally, I performed a wide range of robustness checks such as dropping the outliers, allowing heterogeneous time fixed effect, and using diffrent estimators. See Appendix A for the full set of results. The correlation between the wealthy HtM measure and the local multplier is robust across specifications and the average multiplier is consistently lower than the original estimate.

3.4 Discussion

My empirical analyses have two robust findings. First, there is a sizeable bias in the usual estimates of the average local multiplier, due to the leverage of Bartik-IV and the heterogeneous effects of military spending across states. It is worth noting that Basso and Rachedi (2021) had adopted a nearly the same empirical design as this paper, except that they focused on the share of young workers in each state as a factor of the local multplier, but they did not find a bias in the average multiplier. In Appendix D.2, I use the bias formula shown in Section 3.1 to show that the discrepency is due to the fact that their age-structure factor is *less* correlated with the Bartik shift-share and is a *weak* predictor for the local multiplier. In Appendix E, I calibrate a simple monetary-union TANK model to show that the bias is large enough to matter for inference. The main implication of the calibration exercise is that GHH preference is not needed to match the empirical evidence of local multiplier, thereby reconciling with the findings of Auclert, Bardóczy and Rognlie (2023) that GHH preference in HANK models implies implausibly large national fiscal multplier.

Second, there is substantial variation in the effect of military spending across states and the effect is larger in states with *lower* wealthy HtM measures. The regional heterogeneity in the effect of governement spending has been documented by previous researches using other spending series and methodology. See for example Sarto (2018) and Flynn, Patterson and Sturm (2022). Nonetheless, the negative correlation between the local multplier and the share of wealthy HtM is novel. As shown in the last section, this correlation cannot be explained by other unidimensional demographic factor nor supply-side factor. From a theoretical perspective, the negative correlation is inconsistent with the MPC mechanism championed by the HANK literature. This suggests that there are other crucial mechanisms behind the transmission of military spending. Before delving into the potential mechanisms, I first study what's the implications of the smaller average local multipliers.

4 Implications for aggregate fiscal multipliers

The connection between the estimated local multipliers and the desired aggregate multipliers is far from trivial. As extensively discussed in Chodorow-Reich (2019), there are three major ways in which these two multipliers can diff. First, the local economy is open so the local multipliers would depend on the extent of risk-sharing across regions and the openness of the economy. Second, the diff-in-diff design controls for aggregate policy responses (e.g. monetary policy) and hence the local multiplier should be interpreted as a no-policy-response multiplier. Third, how the taxes are distributed across regions could also affect the local multiplier significantly. All these issues cast doubts on the usefulness of the local multipliers. Nonetheless, Nakamura and Steinsson (2014) argue that we can use the identified local multiplier as a moment to distinguish structural models and then rely on the identified model to infer the aggregate multiplier. In this section, I follow them to perform the structural exercises with TANK models à la Galí et al. (2007) and Bilbiie (2020).

4.1 Monetary union TANK model

The model is an extension of the monetary union model in Nakamura and Steinsson (2014) with a two-agent structure. There are two regions in the model, a home region that corresponds to an average state and a foreign region that represents the rest of the nation. Households' preferences are separable and isoelastic, and there are complete financial markets across regions that only non-HtM households have access to. In the followings, I describe the new ingredients associated with the TA structure.

Hand-to-mouth In each region, a fraction λ of households are assumed to be hand-to-mouth.[11] Their consumption is given by

$$P_{it}C_{it}^H = W_{it}L_{it} + \frac{\delta}{\lambda}D_{it} + T_i - T_{it}^G \qquad \forall i \in \{H, F\}$$

where P_{it} is the price level, C_{it}^H is the consumption of a hand-to-mouth household, W_{it} is the nominal wage, L_{it} is the labor supply, D_{it} is profits, T_i is redistributive transfer, and T_{it}^G is the lump-sum tax to finance government spending. The parameter $\delta \in [0, 1]$ is the fraction of profits that is distributed to the hand-to-mouth households. This is an important parameter as it directly controls the income sensitivity to aggregate income, which is identified in the literature as a crucial determinant of the effect of monetary and fiscal policy in HANK models (Bilbiie 2020, Werning 2015). The redistributive transfer T_i is set to equalize the consumption of the two types of households and is assumed to be invariant to shocks. The financing tax T_{it}^G is specified latter in the government sections. I consider two cases: $\delta = \lambda$ and $\delta = 0$.

Labor market In each region, there is a labor union who maximizes the average welfare of the residents and demands labor from the residents uniformly. That is, the labor supply curve in a region is given by

$$W_{it} = C_{it}^{-\sigma^{-1}} L_{it}^{\nu^{-1}} \qquad \forall i \in \{H, F\}$$

where $C_{it} = \lambda C_{it}^H + (1-\lambda)C_{it}^N$. Note that wages are still flexible, unlike the usual sticky-wage setup à la Erceg et al. (2000). This specification is desirable because it attenuates (but not eliminate) the relationship between individual consumption and labor supply decision which is found to be problematic when studying fiscal multipliers in TANK/HANK models (Broer et al. 2021, Auclert, Bardóczy and Rognlie 2023).[12] Alternative specifications include a worker-capitalist setup where only the consumption of hand-to-mouth households enters the first-order condition (Cantore and Freund 2021) and a competitive market setup where the first-order condition holds for both types of households independently (Galí, López-Salido and Vallés 2007).

[11]The assumption that the home region has the same fraction of HtM households as the foreign region is the right one as we are targeting the average multiplier. Holding the foreign fraction fixed, we can vary the home fraction to compute local multipliers for different levels of HtM.

[12]The individual consumption-labor relationship is not eliminated here simply because the model has only two (representative) agents. In a heterogeneous agent model, the labor union will completely eliminate any individual relationship.

Taxation The government holds a balanced budget every period by imposing uniform lump-sum taxes on all households in the nation:

$$T_{it}^{G} = P_t^H G_{Ht} + P_t^F G_{Ft} \qquad \forall i \in \{H, F\}$$

where P_t^i is the price of region i's goods. Since only non-HtM households have access to financial markets, the tax burden of the HtM households is relevant to the effect of government spending. Here I assume as a benchmark that all households bear the same amount of taxes. Note that local spending is financed nationally.

Appendix E contains all the (log-linearized) model equations and calibration details. The parameters are calibrated as in Nakamura and Steinsson (2014). For the fraction of HtM households λ, I calibrate it to match the predicted value in the census data which is 0.35.

4.2 Simulation results

I simulate the model and estimate the state-CPI-deflated output regression (column 3 of Table 4) with the simulated data. Table 8 reports the results along with the original results in Nakamura and Steinsson (2014). Thanks to the diff-in-diff design, the open economy relative multipliers do not depend on the aggregate monetary policy rule. The benchmark TANK model matches very well to the empirical estimates, while in the set of representative agent models only the one with incomplete markets and federally financed spending can get close. Based on their high estimates of the relative multipliers, Nakamura and Steinsson concluded that models with GHH preferences won the race.[13] When the heterogeneity bias is corrected, the empirical estimates drop substantially and accordingly their conclusion no long holds.[14]

The aggregate multipliers depend strongly on the monetary policy rule. Under Taylor rule, the benchmark TANK model predicts a small multiplier of 0.2, which is similar to the predictions of other RA models. Surprisingly, under constant-nominal-rate policy which should resemble the ZLB scenario, the aggregate multiplier is close to 0. This is because the foreign consumption drops a lot in response to the aggregate demand expansion in the home region initiated by the government spending shock. Overall, the results highlight the

[13]Indeed, in the paper they have hypothesized that TANK models can also match their high estimates. In this regard, my result shows that a model with empirically plausible fraction of HtM households actually fail to generate such a high relative multiplier.

[14]In principle, I should also consider a TANK model with GHH preference. However, the model setup is a bit complicated and I don't have enough time to figure it out. I guess the multipliers will increase a lot. Indeed, Auclert et al. (2023) have shown that in a quantitative HANK model with GHH preference the multiplier will be implausibly large.

Table 8: Fiscal multipliers in monetary union models

	Data	TANK ($\delta = \lambda$)	TANK ($\delta = 0$)	Nakamura and Steinsson				
				Data	Separable	GHH	IM, locally financed	IM, federally financed
Open economy relative multiplier:								
Taylor rule	1.00	1.01	1.27	1.5	0.83	1.42	0.84	0.90
Constant real rate	1.00	1.01	1.27	1.5	0.83	1.42	0.84	0.90
Constant nominal rate	1.00	1.01	1.27	1.5	0.83	1.42	0.84	0.90
Closed economy aggregate multiplier:								
Taylor rule		0.20	0.13		0.2	0.12	0.18	0.18
Constant real rate		1.00	-0.57		1.00	7	0.92	0.93
Constant nominal rate		0.03	0.23		∞	∞	-0.05	-0.04

NOTE. The Nakamura and Steinsson results are from their Table 6-8. The two "Data" columns refer to the reduced-form estimates.

importance of the interregional dynamics in determining the aggregate effect of government spending.

5 Local responses to aggregate shocks and the share of wealthy HtM

In this section, I examine the correlation between the local response to aggregate shocks and the wealthy HtM measure. The purpose of this exercise is twofold. First, if the negative correlation between the local multplier and the wealthy HtM measure is driven by some generic economic factor (for instance, regional heterogeneity in aggregate labor supply), then we should expect to see a similar correlation between the local response and the wealthy HtM measure. Second, the exercise serves as an external test of the relevance of the wealthy HtM measure beyond military spending.

5.1 Data

I build a quarterly panel of the 50 states and D.C. for the period 1975Q1-2010Q4. Since an official measure of quarterly GDP in the state level is only available beginning 2005Q1, I use instead the BEA real labor income per capita as the main outcome variable. I consider three aggregate shocks: 1) innovation to real national GDP growth; 2) the main business cycle (MBC) shock in Angeletos, Collard and Dellas (2020); and 3) the excess bond premium (EBP) shock in Gilchrist and Zakrajšek (2012). The innovation to real national GDP growth is constructed from a quarterly VAR(2) of four variables including real national GDP growth, growth of GDP deflator, 10-year treasury yield, and effective federal funds rate. This reduced-form shock is used to estimate the typical comovement of a state's economy with national economy. The MBC shock is constructed by the authors to maximally explain the business-cycle movement of ten macroeconomic variables such as unemployment and output during 1955Q1-2017Q4. For instance, it can explain about 75% of the volatility of unemployment over the business-cycle frequencies. I use this shock to examine the response of a state's economy to typical business-cycle shocks. Lastly, the EBP shock is a financial shock that triggers a large and persistent drop in investment and stock return, therefore providing another perspective on the regional responses.

5.2 Empirical methodology and results

I use local projection (Jordà 2005) to estimate the correlation between the local response and the wealthy HtM measure. Below is the econometric specification:

$$\sum_{h=0}^{3} \frac{Y_{it+h} - Y_{it+h-1}}{Y_{it-1}} = \beta_0 \log(WHtM_i) \times shock_t + \beta_1 \overline{share_i} \times shock_t + \mathbf{X}'_{it}\gamma + \alpha_i + \alpha_t + \epsilon_{it},$$

where Y_{it} is quarterly real labor income per capita, $\log(WHtM_i)$ is log wealthy HtM measure as constructed before, $shock_t$ is one of the national shocks, $\overline{share_i}$ is average output share of national output, $\mathbf{X}'_{it}$ is a set of controls including house price growth and population growth[15], and the α's are fixed effects. The average output share of national output is computed using the annual GDP during 1966-2006 and is included to control for the relative size of each state's economy. The dependent variable is 1-year cumulative growth of real labor income per capita and the shock is standardized so that the response can be interpreted as 1-year cumulative response to 1-std increase in the shock. The coefficient of interest is β_0, which captures the the correlation between the wealthy HtM measure and the responses. Note that the average effect of the national shock is absorbed by the time fixed effect. To estimate the average effect, I consider another specification where the time fixed effect is replaced with additional aggregate controls including growth of GDP deflator, 10-year treasury yield, and effective federal funds rate.

Table 9 reports the results. Column 1-2 correspond to the results from using GDP growth innovation as the shock. On average, a 1-std innovation to the national GDP growth causes the real labor income per captia in a state to increase by 1.02% after a year. The coefficient on the interaction term with the wealthy HtM measure is positive and significant, meaning that states with more wealthy HtM have *larger* responses.

[15] The hosue price is the "All-Transactions Indexes" from FHFA and the population data is from BEA.

The results from using the MBC shock are similar, with the correlation between the wealthy HtM measure and the size of response slightly larger. Overall, these results imply that states with more wealthy HtM are more procylical, consistent with the findings of Patterson (2021). In contrast, the regional pattern of the responses to EBP shocks is remarkedly different. The estimation results imply that states with more wealthy HtM are *less* responsive to EBP shocks, though the associated coefficient is only significant at 10% level. Note that in all specifications, states with higher output shares are always more responsive to aggregate shocks, so the difference between the two regional patterns cannot be attributed to size. In Table A.5, I show that the results are robust to adding more lags of the controls.

Taking stock of the evidences, states with more wealthy HtM are more procyclical in general but less responsive to financial shocks. Recall that local military spending is less stimulative in states with more wealthy HtM. What mechanism can possibly explain the differences in the regional patterns? We can rule out factors which affect the transmission of any aggregate shocks in the same way, for instance labor supply elasticity and preference. The "matching multplier" mechanism of Patterson (2021) can explain the differences if financial shocks and military spendings target towards rich and high-income households. Financial factors may also play a role. Juarros (2021) finds that local military spending increases the investment of small firms in the area and relaxes their borrowing constraints, suggesting that financial factors can intermediate the transmission of military spending. Overall, the empirical evidences imply that the nature of the shock matters for the heterogeneity in local responses and that the transmission of military spending can be very different from generic expansionary shocks.

6 Micro evidence on the transmission of military procurement spending

In this section, I study the micro-details of military procurement spending and examine if and how it matters for the propagation of military spending to the whole economy.

6.1 Data

I build a military procurement contract dataset for the period 2001-2019 using the Department of Defense (DOD) contract database at USAspending.gov. The raw data are based on DD-350 and DD-1057 military procurement forms and the database covers the universe of

Table 9: Heterogeneous local responses to aggregate shocks

	(1)	(2)	(3)	(4)	(5)	(6)
Shock:	INV	INV	MBC	MBC	EBP	EBP
β_0	0.0131***	0.0131***	-0.0202***	-0.0204***	0.0037*	0.0042*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)
β_1	0.0344**	0.0336**	-0.0339*	-0.0338*	-0.0316***	-0.0307***
	(0.013)	(0.013)	(0.019)	(0.019)	(0.008)	(0.008)
$shock_t$		1.0197***		-1.2278***		-0.2003***
		(0.037)		(0.055)		(0.026)
$\beta_0/\beta(shock_t)$		0.0128		0.0166		-0.0209
Observations	7,000	7,000	7,000	7,000	7,000	7,000
R-squared	0.444	0.161	0.445	0.180	0.443	0.087
Year	1975Q1-2010Q4	1975Q1-2010Q4	1975Q1-2010Q4	1975Q1-2010Q4	1975Q1-2010Q4	1975Q1-2010Q4
Control	TWFE	state FE + agg.	TWFE	state FE + agg.	TWFE	state FE + agg.

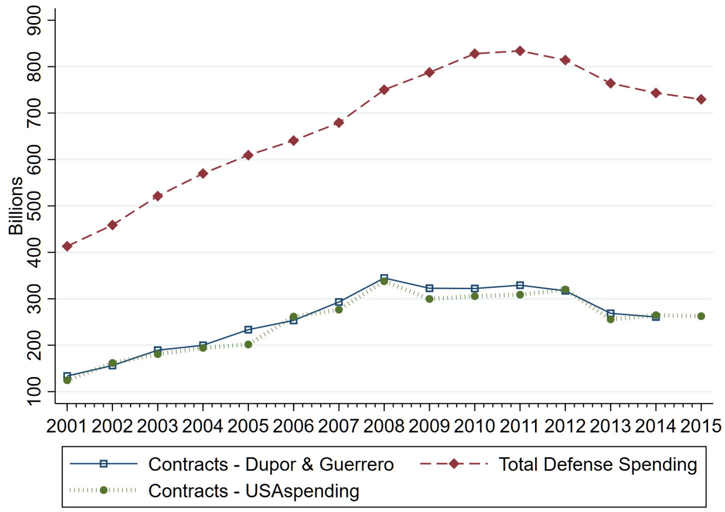
NOTE. The unit of the coefficient is percentage point. Column (2), (4), and (6) control for inflation, 10-year treasury yield, and effective federal funds rate instead of time FE. Standard errors are clustered by state. *** p<0.01, ** p<0.05, * p<0.1

DOD contracts signed from fiscal year 2001 onwards.[16] Each observation is a unique contract between the DOD and a prime contractor. For each contract, we observe the obligated funds, the duration of the contract, and all subsequent modifications since the sign date. We also observe the name, industry (6-digit NAICS), and tax ID of the contractor as well as the location of performance at zip-code level. Following the procedures in Demyanyk, Loutskina and Murphy (2019), I construct the military spending variable by evenly distribute the obligated funds over the duration of the contract. To validate the result, I aggregate the contract-level spendings to the national level and compare it with military spending series from different sources. Figure 4 is the time-series plot for year 2001-2015. The red line is the total military spending series from BEA and the blue line is the military contract spending series constructed by Dupor and Guerrero (2017) using official reports published by the Directorate for Information Operations and Reports of the DOD. We can see that at the national level the contract spending from the DOD contract database tracks closely the official series, alleviating concerns of under reporting.[17]

[16]The state-level military procurement spending data constructed by Nakamura and Steinsson (2014) is also based on DD-350 and DD-1057 forms.

[17]There is an obvious gap in 2005.

Figure 4: National Military Spending



6.2 Composition of military procurement spending

The industry information on the contract allows us to look into the composition of military procurement spending.[18] For each industry, I compute the average share of total military procurement spending on the industry over 2001-2019. Table 10 presents the results for the five 3-digit NAICS industries with the highest shares. Not surprisingly, a large fraction of the spending go towards Manufacturing of Transportation Equipments (336) which includes the manufacturing of aircrafts and tanks. In particular, the sub-industry, Manufacturing of Aircraft and Missile (3364), per se accounts for 20% of the total spending. Another industry that accept a large fraction of the spending is Professional, Scientific, and Technical Services (541), which accounts for 25% of the total spending. Note that the top two 3-digit NAICS industries together already account for more than half of the military procurement spending. For comparison, during the same period, the average value-added share of U.S. GDP of the two industries are 1.6% (336) and 7% (541), respectively. The composition of military procurement spending is therefore very different from the composition of the U.S. economy. Bouakez, Rachedi and Santoro (2022) show that the input-output network can greatly affect the aggregate effect of governement spending, so the special composition of military procurement spending can be of first-order importance.

From the aggregate demand perspective, it is also important who are the households that benefit directly from military procurement spending. I use the ACS (2001-2019) to explore the demographic composition of each industry. I assign each household to an industry based

[18] Auerbach, Gorodnichenko and Murphy (2020) also documented the composition of military procurement spending using the DOD contract dataset.

Table 10: Composition of Military Procurement Spending

Top five 3-digit NAICS		Share
336	**Manufacturing (Transportation equipment)**	**.288**
	Aircraft, Missile (3364)	.204
	Ship (3366)	.048
	Tank (3369)	.022
541	**Professional, Scientific, and Technical Services**	**.254**
	Engineering Service (5413)	.093
	R & D (5417)	.081
	Computer Programming (5415)	.043
334	Manufacturing (Computer and electronic products)	.074
236	Construction (industrial building)	.040
524	Insurance	.031

Table 11: Demographic Composition of Major Industries

NAICS		Share	WHtM	PHtM	College	Homeowner	(median) labor earnings
541	Professional, Scientific, and Technical Services	.254	.165	.055	.667	.673	61165
3364	Manufacturing (Aircraft, Missile)	.204	.187	.046	.467	.782	63602
334	Manufacturing (Computer and electronic products)	.074	.178	.060	.517	.734	62577
23	Construction	.071	.226	.138	.129	.641	36465
3366	Manufacturing (Ship)	.048	.231	.092	.185	.709	42900
524	Insurance	.031	.206	.068	.481	.716	52320
3369	Manufacturing (Tank)	.022	.223	.096	.241	.695	41958
National avg:			.210	.130	.333	.615	36601

NOTE. All statistics are computed over year 2001 - 2019. The demographic refers to the head of the household. Labor earnings is in 1999 US dollars.

on the industry of the household's head. To relate to the empirical results in Section 3, I also impute the share of HtM in each industry using the same imputation approach. Since the industry of the manufacturing of aircraft and missile (3364) is special, I decompose its mother industry (336) into 4-digit sub-industries. The other industries are all in 3-digit level unless the ACS does not provide 3-digit NAICS information. Table 11 shows the results. The demographic composition of the top 3 industries (541, 3364, 334) is remarkedly differently from the rest. Households in these industries are more educated, likely own a house, and have higher labor earnings. As a result, they are also less likely to be HtM. This is not a surprising finding as these industries are high-tech and high-skilled. What is surprising and economically interesting is that a majority of the military procurement spending tilts towards industries with high-skilled, high-income, and rich households who are not HtM and have low MPC. This mismatch can dampen the consumption channel and attenuate the aggregate effects of governement spending.

6.3 Local industry response to military spending shock

Does the effect of military spending on different industry differ? To answer this question, I complement the contract dataset with local industry labor statistics from the Quarterly Census of Employment and Wages (QCEW). The QCEW is maintained by the BLS and provides quarterly employment and wages statistics at different geographic and industry level. Due to data availability, I focus on county-3-digit-NAICS level.[19] To obtain a balanced panel, I drop county-industry pairs with incomplete records.[20] The resulting panel consists of 4,579 county-industry pair over the year 2001-2019. To estimate the effect of military spending on the economic activity of a industry, I consider the following econometric specification

$$\frac{Y_{il,t} - Y_{il,t-2}}{Y_{il,t-2}} = \beta \frac{G_{il,t} - G_{il,t-2}}{Y_{il,t-2}} + \alpha_{il} + \alpha_t + \epsilon_{il,t}$$

where $Y_{il,t}$ is either total wages, average weekly wages (per worker), or average monthly employment of industry i in county l in year t, $G_{il,t}$ is military procurement spending, α_{il} is county-industry fixed effect, and α_t is year fixed effect. Following Auerbach, Gorodnichenko and Murphy (2020), I instrument military spending with Bartik IV, $s_{il} \times \frac{G_t - G_{t-2}}{Y_{l,t-2}}$, where the Bartik shift-share $s_{il} := \sum_t G_{il,t} / \sum_t G_t$ is the average spending share of national spending, G_t is national military procurement spending, and $Y_{l,t-2}$ is total wages of the corresponding county.

Table 12 reports the baseline results. Standard errors are clustered by state. Consistent with the state-level findings, military spendings stimulate the economic activity of a industry, increasing both employment and wages. On average, one-dollar increase in the spending causes about 28 cent increase in total wages. Assuming a labor share of 2/3, the implied output multiplier is around .45.

Based on the findings in the last section, I divide the sample into two groups, the top-3 industries (NAICS 336, 541, 334) and the rests, and reestimate the effect on each subsample. Table 13 reports the results. Comparing the coefficients across the columns, we can see that military procurement spending is more stimulative in the top-3 industries than in the others.

[19] The finest level available is county-6-digit-NAICS but in this level most of the data are not disclosed. Indeed, even in the county-3-digit-NAICS level, about 20% of the county-industry-year observations are missing.

[20] In Appendix A, I report results from using all nonmissing observations.

Table 12: Effect of Military Procurement Spending on Local Industry

	(1) Wages	(2) Avg. weekly wages	(3) Employment
β	0.276***	0.067***	0.190***
	(0.049)	(0.014)	(0.032)
Observations	77,843	77,843	77,843
Year	2001-2019	2001-2019	2001-2019
County-Industry	4579	4579	4579
F stat	38.70	38.70	38.70

NOTE. Standard errors are clusted by state. *** p<0.01, ** p<0.05, * p<0.1

Table 13: Effect on Major Industries vs. Others

	(1) Wages	(2) Avg. weekly wages	(3) Employment	(4) Wages	(5) Avg. weekly wages	(6) Employment
β	0.380***	0.083***	0.266***	0.135*	0.040*	0.087**
	(0.041)	(0.011)	(0.032)	(0.070)	(0.022)	(0.043)
Observations	15,878	15,878	15,878	61,965	61,965	61,965
Year	2001-2019	2001-2019	2001-2019	2001-2019	2001-2019	2001-2019
County-Industry	934	934	934	3645	3645	3645
Major industry	YES	YES	YES	NO	NO	NO
F stat	38.82	38.82	38.82	4.97	4.97	4.97

NOTE. Major industries are NAICS 336, 541, and 334. Standard errors are clusted by state. *** p<0.01, ** p<0.05, * p<0.1

Importantly, the first-stage F statistic for the minor industries is small, while the F statistic for the top-3 industries is large and close to the F statistic for the full sample reported in Table 12. This means that most of the variation in the Bartik IV is due to the variation in the spending on the top-3 industries. If the effect and transmission mechanism of government spending on these industries are special, then the fiscal multplier identifed using Bartik-IV and military spending might not be externally valid.

Finally, I examine the effects of different spendings on the whole economy. The unit of analysis is county and I distinguish the military spending on the top-3 industries from the others. As a result, the sample contains only counties that accept both types of military

Table 14: Heterogeneous Aggregate Effects of Different Spending

	(1) Wages	(2) Avg. weekly wages	(3) Employment
β_0	0.450***	0.238***	0.194***
	(0.086)	(0.043)	(0.069)
β_1	0.996*	0.493**	0.474
	(0.562)	(0.207)	(0.348)
Observations	7,650	7,650	7,650
Year	2001-2019	2001-2019	2001-2019
County	450	450	450
F stat	52.34	52.34	52.34

NOTE. Standard errors are clusted by state. *** p<0.01, ** p<0.05, * p<0.1

spendings. The specification is as follows

$$\frac{Y_{l,t} - Y_{l,t-2}}{Y_{l,t-2}} = \beta_0 \frac{G^0_{l,t} - G^0_{l,t-2}}{Y_{l,t-2}} + \beta_1 \frac{G^1_{l,t} - G^1_{l,t-2}}{Y_{l,t-2}} + \alpha_l + \alpha_t + \epsilon_{l,t}$$

where $Y_{l,t}$ is one of the outcome variables at the county level, $G^0_{l,t}$ ($G^1_{l,t}$) is the military spending on the major (minor) industries in county l and year t, α_l is county fixed effect, and α_t is year fixed effect. I instrument each spending variable using Bartik-IV constructed in the same way as before. Table 14 presents the results. The point estimates imply that the overall effect of spending on the minor industries is *stronger*, though the effects are less accurately estimated as indicated by the relatively large standard errors. Overall, the evidences support the hypothesis that military spending on the top-3 industries have different effects on the aggregate economy. What drives the differences merit further studies.

7 Conclusion

Regional variation opens up new opportunities for identifying causal effects of governement policy, but regional heterogeneity can affect the identification if not handled properly. In this paper, I show that the effects of military spending vary substantially across states in the US. Specifically, the local multplier is larger in states with less wealthy HtM. This heterogeneous effect significantly bias previous estimates of the average effect. On the other hand, I argue that the estimated cross-regional pattern of the local multpliers is a specific feature of military spending. Using contract-level data, I document that the composition of military spending is highly skewed – three industries already account for 60% of the spending. I further show that governement spending is particularly stimulative in these three industries but spendings on these industries are less effective in stimulating the aggregate economy, suggesting that the composition of spending matters for the transmission. Interpretation of the fical multplier identifed using regional variation in military spending therefore requires a structural model that takes into account 1) regional heterogeneity; and 2) the unusal composition of military spending. I leave this task to the future project.

References

Aguiar, Mark A, Mark Bils, and Corina Boar (2021) "Who are the Hand-to-Mouth?"Technical report, National Bureau of Economic Research.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas (2020) "Business-cycle anatomy," *American Economic Review*, 110 (10), 3030–70.

Auclert, Adrien, Bence Bardóczy, and Matthew Rognlie (2023) "MPCs, MPEs, and multipliers: A trilemma for New Keynesian models," *Review of Economics and Statistics*, 105 (3), 700–712.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub (2018) "The intertemporal keynesian cross,"Technical report, National Bureau of Economic Research.

Auerbach, Alan, Yuriy Gorodnichenko, and Daniel Murphy (2020) "Local fiscal multipliers and fiscal spillovers in the USA," *IMF Economic Review*, 68 (1), 195–229.

Basso, Henrique S. and Omar Rachedi (2021) "The Young, the Old, and the Government: Demographics and Fiscal Multipliers," *American Economic Journal: Macroeconomics*, 13 (4), 110–41.

Bilbiie, Florin O (2020) "The new Keynesian cross," *Journal of Monetary Economics*, 114, 90–108.

Bouakez, Hafedh, Omar Rachedi, and Emiliano Santoro (2022) "The government spending multiplier in a multi-sector economy," *American Economic Journal: Macroeconomics*, forthcoming.

Brinca, Pedro, Hans A Holter, Per Krusell, and Laurence Malafry (2016) "Fiscal multipliers in the 21st century," *Journal of Monetary Economics*, 77, 53–69.

Broer, Tobias, Per Krusell, and Erik Öberg (2021) "Fiscal Multipliers: A Heterogenous-Agent Perspective,"Technical report, National Bureau of Economic Research.

Cantore, Cristiano and Lukas B Freund (2021) "Workers, capitalists, and the government: fiscal policy and income (re) distribution," *Journal of Monetary Economics*, 119, 58–74.

Chodorow-Reich, Gabriel (2019) "Geographic cross-sectional fiscal spending multipliers: What have we learned?" *American Economic Journal: Economic Policy*, 11 (2), 1–34.

Demyanyk, Yuliya, Elena Loutskina, and Daniel Murphy (2019) "Fiscal stimulus and consumer debt," *Review of Economics and Statistics*, 101 (4), 728–741.

Dupor, Bill and Rodrigo Guerrero (2017) "Local and aggregate fiscal policy multipliers," *Journal of Monetary Economics*, 92, 16–30.

Erceg, Christopher J, Dale W Henderson, and Andrew T Levin (2000) "Optimal monetary policy with staggered wage and price contracts," *Journal of monetary Economics*, 46 (2), 281–313.

Farhi, Emmanuel and Iván Werning (2016) "Fiscal multipliers: Liquidity traps and currency unions," in *Handbook of macroeconomics*, 2, 2417–2492: Elsevier.

Flynn, Joel P, Christina Patterson, and John Sturm (2022) "Fiscal policy in a networked economy,"Technical report, National Bureau of Economic Research.

Galí, Jordi, J David López-Salido, and Javier Vallés (2007) "Understanding the effects of government spending on consumption," *Journal of the european economic association*, 5 (1), 227–270.

Gilchrist, Simon and Egon Zakrajšek (2012) "Credit spreads and business cycle fluctuations," *American economic review*, 102 (4), 1692–1720.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift (2020) "Bartik instruments: What, when, why, and how," *American Economic Review*, 110 (8), 2586–2624.

Jordà, Òscar (2005) "Estimation and inference of impulse responses by local projections," *American economic review*, 95 (1), 161–182.

Juarros, Pedro (2021) "Fiscal stimulus, credit frictions and the amplification effects of small firms,"Technical report.

Kaplan, Greg, Giovanni L Violante, and Justin Weidner (2014) "The wealthy hand-to-mouth,"Technical report, National Bureau of Economic Research.

Nakamura, Emi and Jon Steinsson (2014) "Fiscal stimulus in a monetary union: Evidence from US regions," *American Economic Review*, 104 (3), 753–92.

Nakamura, Emi and Jón Steinsson (2018) "Identification in macroeconomics," *Journal of Economic Perspectives*, 32 (3), 59–86.

Orchard, Jacob, Valerie A Ramey, and Johannes F Wieland (2023) "Micro MPCs and macro counterfactuals: the case of the 2008 rebates,"Technical report, National Bureau of Economic Research.

Parker, Jonathan A, Nicholas S Souleles, David S Johnson, and Robert McClelland (2013) "Consumer spending and the economic stimulus payments of 2008," *American Economic Review*, 103 (6), 2530–2553.

Patterson, Christina (2021) "The matching multiplier and the amplification of recessions," *Unpublished Manuscript, University of Chicago, Booth School of Business.*

Ramey, Valerie A (2011) "Can government purchases stimulate the economy?" *Journal of Economic Literature*, 49 (3), 673–85.

Sarto, Andres (2018) "Recovering Macro Elasticities from Regional Data," *Unpublished manuscript, NYU Stern.*

Sun, Liyang and Jesse M Shapiro (2022) "A linear panel model with heterogeneous coefficients and variation in exposure," *Journal of Economic Perspectives*, 36 (4), 193–204.

Werning, Iván (2015) "Incomplete markets and aggregate demand,"Technical report, National Bureau of Economic Research.

Wooldridge, Jeffrey M (2015) "Control function methods in applied econometrics," *Journal of Human Resources*, 50 (2), 420–445.

Appendix A Supplementary tables and figures

Table A.1: Robustness: Heterogeneous Time-FE

	(1) Output	(2) Output defl. state CPI	(3) Employment	(4) CPI
$\frac{G_{it}-G_{it-2}}{Y_{it-2}}$	1.06***	0.93***	1.05***	0.21
	(0.36)	(0.32)	(0.24)	(0.18)
$\log(WHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$	-0.05	-0.06*	-0.01	0.02
	(0.05)	(0.03)	(0.02)	(0.02)
$\log(WHtM_i) \times \gamma_t$	YES	YES	YES	YES
Observations	1,989	1,989	1,989	1,763

NOTE. The interactive fixed effect $\log(WHtM_i) \times \gamma_t$ is included.

Table A.2: Robustness: Dropping Outliers

	(1) Output	(2) Output defl. state CPI	(3) Employment	(4) CPI
$\frac{G_{it}-G_{it-2}}{Y_{it-2}}$	1.17***	0.99***	0.99***	0.17
	(0.32)	(0.29)	(0.24)	(0.15)
$\log(WHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$	-0.03	-0.06***	-0.05***	0.05***
	(0.02)	(0.02)	(0.02)	(0.01)
Observations	1,872	1,872	1,872	1,680

NOTE. State with the highest share of WHtM (i.e. MS), state with the lowest share of WHtM (i.e. AK), and DC are dropped.

Table A.3: Robustness: Heterogeneous Time-FE in the First Stage

	(1) Output	(2) Output defl. state CPI	(3) Employment	(4) CPI
$\frac{G_{it}-G_{it-2}}{Y_{it-2}}$	0.95***	0.91***	0.94***	-0.12
	(0.26)	(0.31)	(0.26)	(0.22)
$\log(WHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$	-0.05	-0.09*	-0.06***	0.06***
	(0.04)	(0.05)	(0.02)	(0.01)
Observations	1,989	1,989	1,989	1,763

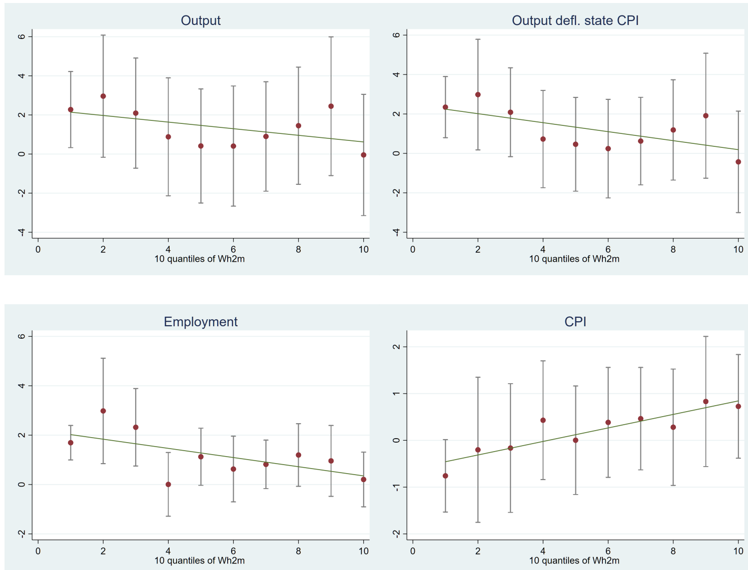
NOTE. The interactive fixed effect $\log(WHtM_i) \times \gamma_t$ is included *only* in the first-statge.

Table A.4: Robustness: Control Function Approach

	(1) Output	(2) Output defl. state CPI	(3) Employment	(4) CPI
$\frac{G_{it}-G_{it-2}}{Y_{it-2}}$	1.03**	0.90**	1.06***	-0.00
	(0.41)	(0.35)	(0.26)	(0.12)
$\log(WHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$	-0.08	-0.09**	-0.05***	0.04***
	(0.05)	(0.04)	(0.02)	(0.01)
$\log(PHtM_i) \times \frac{G_{it}-G_{it-2}}{Y_{it-2}}$	-0.02	-0.01	0.00	0.00
	(0.02)	(0.02)	(0.01)	(0.01)
Observations	1,989	1,989	1,989	1,763

NOTE. The model is estimated using the control function approach in Wooldridge (2015). Standard errors are clustered by state BUT NOT ADJUSTED FOR CONTROL FUNCTION!

Figure 5: Grouped Fixed Effect



NOTE. States are divided into 10 decile groups based on WHtM measures.

Table A.5: Heterogeneous local responses to aggregate shocks (control for two lags)

	(1)	(2)	(3)	(4)	(5)	(6)
Shock:	INV	INV	MBC	MBC	EBP	EBP
β_0	0.0120**	0.0121***	-0.0196***	-0.0203***	0.0035*	0.0034
	(0.005)	(0.004)	(0.005)	(0.005)	(0.002)	(0.002)
β_1	0.0303**	0.0278*	-0.0283	-0.0258	-0.0314***	-0.0307***
	(0.014)	(0.015)	(0.020)	(0.021)	(0.009)	(0.010)
$shock_t$		0.8987***		-1.1419***		-0.1308***
		(0.040)		(0.058)		(0.025)
$\beta_0/\beta(shock_t)$		0.0135		0.0178		-0.0259
Observations	6,900	6,900	6,900	6,900	6,900	6,900
R-squared	0.461	0.226	0.462	0.240	0.460	0.178
Year	1975Q1-2010Q4	1975Q1-2010Q4	1975Q1-2010Q4	1975Q1-2010Q4	1975Q1-2010Q4	1975Q1-2010Q4
Control	TWFE	state FE + agg.	TWFE	state FE + agg.	TWFE	state FE + agg.

NOTE. Two lags of each of the control variables are added.

Table A.6: Effect of Military Procurement Spending on Local Industry (all nonmissing obs.)

	(1) Wages	(2) Avg. weekly wages	(3) Employment
β	0.197***	0.052***	0.166***
	(0.067)	(0.012)	(0.046)
Observations	266,701	266,701	266,701
Year	2001-2019	2001-2019	2001-2019
County-Industry	28534	28534	28534
F stat	13.20	13.20	13.20

Note. All county-industry-year observations with nonmissing data are used in the estimation. Standard errors are clusted by state. *** p<0.01, ** p<0.05, * p<0.1

Table A.7: Effect on Major Industries vs. Others (all nonmissing obs.)

	(1) Wages	(2) Avg. weekly wages	(3) Employment	(4) Wages	(5) Avg. weekly wages	(6) Employment
β	0.300***	0.055***	0.221***	0.136	0.045**	0.135
	(0.051)	(0.015)	(0.037)	(0.128)	(0.021)	(0.100)
Observations	29,009	29,009	29,009	237,692	237,692	237,692
Year	2001-2019	2001-2019	2001-2019	2001-2019	2001-2019	2001-2019
County-Industry	2597	2597	2597	25937	25937	25937
Major industry	YES	YES	YES	NO	NO	NO
F stat	20.76	20.76	20.76	2.74	2.74	2.74

Note. All county-industry-year observations with nonmissing data are used in the estimation. Major industries are NAICS 336, 541, and 334. Standard errors are clusted by state. *** p<0.01, ** p<0.05, * p<0.1

Appendix B Definition of wealth and income

All definitions follow Kaplan et al. (2014). See their section III.A for detail discussions.

Income The definition is chosen to include all labor income plus any regular government transfers. Importantly, interest, dividend, and other capital income are excluded.

$$\begin{aligned}\text{Income} = &\text{ gross wages and salaries} + \text{self-employment income}\\ &+ \text{regular private transfers (e.g. child support)} + \text{public transfers (e.g. UI benefits)}\\ &+ \text{other regular income (excluding investment income)}\end{aligned}$$

Liquid wealth

$$\begin{aligned}\text{Liquid asset} &= \text{checking, saving, money market, and call accounts}\\ &\quad + \text{directly held mutual funds, stocks, corporate bonds, and government bonds}\\ \text{Liquid debt} &= \text{sum of all credit card balances that accrue interest}\\ \text{Net liquid wealth} &= \text{Liquid asset - Liquid debt}\end{aligned}$$

Illiquid wealth

$$\begin{aligned}\text{Illiquid asset} &= \text{value of housing, residential and nonresidential real estate}\\ &\quad + \text{private retirement accounts (e.g. 401(k)s)} + \text{life insurance}\\ &\quad + \text{certificates of deposits} + \text{saving bonds}\\ \text{Illiquid debt} &= \text{mortgages and home equity loans}\\ \text{Net Illiquid wealth} &= \text{Illiquid asset - Illiquid debt}\end{aligned}$$

Appendix C Accuracy analyses of the HtM prediction model

C.1 Bootstrap results

To examine the accuracy of the prediction model, I perform a bootstrap exercise where for each year I repeatedly draw a 5% sample and compute the prediction error for 100 times. Table C.1a reports summary statistics of the bootstrap sample. The bootstrap sample features enough variation in the composition of HtM households to be a good testing ground.[21] The prediction model performs well: the mean prediction error is zero and the standard error of the prediction errors is roughly 0.02. The good performance is not guaranteed as the sample is not drawn from the entire population but the sub-population clustered by year. Table C.1b presents the results of regressing the true values over the predicted values. We can see that the model on average over-predicts the fraction of wealthy HtM households but under-predicts the fraction of poor HtM households. Most importantly, both coefficients are positive so the cross-sectional order in fractions of HtM households are preserved by the prediction model. Overall, the bootstrap results are very positive.

Table C.1: Bootstrap results

(a) Summary statistics

VARIABLES	N	mean	sd	min	max
W-HtM	800	0.204	0.023	0.124	0.289
P-HtM	800	0.135	0.019	0.085	0.210
Error on W-HtM	800	0.000	0.022	-0.071	0.065
Error on P-HtM	800	0.000	0.014	-0.060	0.035

(b) Regression

VARIABLES	W-HtM	P-HtM
Prediction	0.931	1.256
	(0.0971)	(0.0451)
Constant	0.0140	-0.0344
	(0.0199)	(0.0060)
Observations	800	800
R-squared	0.107	0.487

NOTE. Each observation is a 5% random sample from one of the SCF survey year during 1989-2010.

[21] Since the ultimate goal is to predict the fraction of HtM households in the state sample, ideally we should draw sample from a distribution that resembles well the state sample.

Appendix D Bias decomposition

D.1 Estimand of the Bartik-TWFE 2SLS estimator

Here I derive the formula for the estimand of Bartik-TWFE 2SLS estimator presented in the main text. Recall that the structural model is

$$\begin{aligned} Y_{it} &= \beta_i G_{it} + \alpha_i + \alpha_t + \varepsilon_{it} \\ G_{it} &= \gamma_i G_t + \alpha_i^G + \alpha_t^G + \nu_{it} \end{aligned}$$

The reduced form equation for Y_{it} is

$$Y_{it} = \beta_i \gamma_i G_t + (\alpha_i + \beta_i \alpha_i^G) + (\alpha_t + \beta_i \alpha_t^G) + (\varepsilon_{it} + \beta_i \nu_{it})$$

First, take within-unit transformation:

$$\hat{Y}_{it} = \beta_i \gamma_i \hat{G}_t + (\hat{\alpha}_t + \beta_i \hat{\alpha}_t^G) + (\hat{\varepsilon}_{it} + \beta_i \hat{\nu}_{it})$$

where $\hat{X}_{it} := X_{it} - \frac{1}{T}\sum_t X_{it}$ denotes the within-unit deviation. Next, take within-time transformation:

$$Y_{it}^* = \widetilde{\beta_i \gamma_i} \hat{G}_t + \tilde{\beta}_i \hat{\alpha}_t^G + (\varepsilon_{it}^* + \widetilde{\beta_i \hat{\nu}_{it}})$$

where $\tilde{X}_{it} := X_{it} - \frac{1}{N}\sum_j X_{jt}$ denotes the within-time deviation and $X_{it}^* := \tilde{\hat{X}}_{it}$ denotes the product of applying both transformation. Empirically, a researcher instead estimates

$$Y_{it}^* = \beta \tilde{\gamma}_i \hat{G}_t + (\varepsilon_{it}^* + \beta \hat{\nu}_{it})$$

That is, regress Y_{it}^* over $\tilde{\gamma}_i \hat{G}_t$. Along with the usual regularity assumptions, we make the following assumptions:

1. $\mathbb{E}[\gamma_i G_t \varepsilon_{it} \mid \alpha_{it}] = 0$

2. $\mathbb{E}[\beta_i \nu_{it} \mid \alpha_{it}] = 0$

where $\alpha_{it} := (\alpha_i, \alpha_t, \alpha_i^G, \alpha_t^G)'$ is the random vector of fixed effects. Assumption 1 is the classic exclusion restriction of IV and is also the standard identifying assumption made in the literature (Nakamura and Steinsson 2014). Assumption 2 is an exogeneity assumption on the multipliers. It is satisfied if the idiosyncratic movement of the local governement spending is independent of the local multplier. Given these assumptions, we can apply the standard OLS formula and obtain the asymptotic result:

$$\beta^{TWFE} = \frac{\mathbb{E}[\tilde{\gamma}_i \hat{G}_t \widetilde{\beta_i \gamma_i \hat{G}_t}]}{\mathbb{E}[(\tilde{\gamma}_i \hat{G}_t)^2]} + \frac{\mathbb{E}[\tilde{\gamma}_i \hat{G}_t \tilde{\beta}_i \hat{\alpha}_t^G]}{\mathbb{E}[(\tilde{\gamma}_i \hat{G}_t)^2]}$$

Observe that $G_t \perp (\beta_i, \gamma_i)$. Thus, the formula further simplifies to

$$\begin{aligned}
\beta^{TWFE} &= \frac{\mathbb{E}[\tilde{\gamma}_i \cdot \widetilde{\beta_i \gamma_i}]}{\mathbb{E}[\tilde{\gamma}_i{}^2]} + \frac{\mathrm{Cov}(\beta_i, \gamma_i)}{\mathbb{E}[\tilde{\gamma}_i{}^2]} \cdot \frac{\mathbb{E}[\hat{G}_t \hat{\alpha}_t^G]}{\mathbb{E}[\hat{G}_t^2]} \\
&= \mathbb{E}[\beta_i] + \mathbb{E}\left[\frac{\tilde{\gamma}_i{}^2}{\mathbb{E}[\tilde{\gamma}_i{}^2]} \cdot \tilde{\beta}_i\right] + \frac{\mathrm{Cov}(\beta_i, \gamma_i)}{\mathbb{E}[\tilde{\gamma}_i{}^2]} \cdot \left(\mathbb{E}[\gamma_i] + \frac{\mathbb{E}[\hat{G}_t \hat{\alpha}_t^G]}{\mathbb{E}[\hat{G}_t^2]}\right)
\end{aligned}$$

If $\beta_i \perp \gamma_i$, the estimand is exactly the average multiplier $\mathbb{E}[\beta_i]$. If instead $\mathrm{Cov}(\beta_i, \gamma_i) = 0$, then the estimand is a convex combination of the multiplier with weights $\tilde{\gamma}_i{}^2 / \mathbb{E}[\tilde{\gamma}_i{}^2]$.

D.2 Bias accounting

Suppose the multiplier has the factor structure.

$$\beta_i = \mathbb{E}[\beta_i] + \delta_Z Z_i$$

where Z_i is mean-zero and has unit standard deviation. Using the decomposition formula, we can write the bias as

$$\beta^{TWFE} - \mathbb{E}[\beta_i] = \delta_Z \cdot \left\{ \mathbb{E}\left[\frac{\tilde{\gamma}_i{}^2}{\mathbb{E}[\tilde{\gamma}_i{}^2]} \cdot Z_i\right] + \frac{\mathrm{Cov}(Z_i, \gamma_i)}{\mathbb{E}[\tilde{\gamma}_i{}^2]} \cdot \left(\mathbb{E}[\gamma_i] + \frac{\mathbb{E}[\hat{G}_t \hat{\alpha}_t^G]}{\mathbb{E}[\hat{G}_t^2]}\right) \right\}$$

Given data on the factor Z_i, we can estimate the terms in the bracket with plug-in estimator. The coefficient δ_Z can be estimated from the panel regression in the main text. I perform this decomposition exercise for the HtM measures and the age-structure factor of Basso and Rachedi (2021). Table D.1 reports the results.

Table D.1: Bias Accouting

	(1) $\log(WHtM_i)$	(2) $\log(PHtM_i)$	(3) log(Basso-Rachedi$_i$)
$\mathbb{E}[\tilde{\gamma_i}^2 Z_i]/\mathbb{E}[\tilde{\gamma_i}^2]$	-0.1262	-0.1824	-0.2070
$\text{Cov}(Z_i, \gamma_i)/\mathbb{E}[\tilde{\gamma_i}^2]$	-0.3312	-0.0902	-0.1587
$\mathbb{E}[\gamma_i] + \mathbb{E}[\hat{G}_t \hat{\alpha}_t^G]/\mathbb{E}[\hat{G}_t^2]$	0.8507	0.8507	0.8507
Bias ($\delta_Z = 1$)	-0.4080	-0.2591	-0.3420
δ_Z	-0.9419	-0.1629	0.2018
Bias	0.3843	0.0422	-0.069

Appendix E Monetary Union TANK Model

E.1 Model equations

Demand block

$$
\begin{aligned}
c_t^N &= -\sigma(r_t^n - \mathbb{E}_t\pi_{t+1}) + \mathbb{E}_t c_{t+1}^N && \text{(Euler)}\\
c_t^N - c_t^{N*} &= \sigma q_t && \text{(Backus-Smith)}\\
q_t &= p_t^* - p_t && \text{(real exchange rate)}\\
c_t^K + p_t &= \frac{1}{\bar{C}}\left[\left(1-\frac{\delta}{\lambda}\right)\frac{\theta-1}{\theta}a(w_t + p_t + \ell_t) + \frac{\delta}{\lambda}(p_{Ht} + y_{Ht}) - T_t^G\right] && \text{(home HtM)}\\
c_t^{K*} + p_t^* &= \frac{1}{\bar{C}}\left[\left(1-\frac{\delta}{\lambda}\right)\frac{\theta-1}{\theta}a(w_t^* + p_t^* + \ell_t^*) + \frac{\delta}{\lambda}(p_{Ft} + y_{Ft}) - T_t^G\right] && \text{(foreign HtM)}\\
c_t &= \lambda c_t^K + (1-\lambda)c_t^N && \text{(home consumption)}\\
c_t^* &= \lambda c_t^{K*} + (1-\lambda)c_t^{N*} && \text{(foreign consumption)}
\end{aligned}
$$

Supply block

$$
\begin{aligned}
\pi_{Ht} &= \kappa\zeta(\sigma^{-1}c_t + \psi_\nu y_{Ht} + p_t - p_{Ht}) + \beta\mathbb{E}_t\pi_{Ht+1} && \text{(home NKPC)}\\
\pi_{Ft} &= \kappa\zeta(\sigma^{-1}c_t^* + \psi_\nu y_{Ft} + p_t^* - p_{Ft}) + \beta\mathbb{E}_t\pi_{Ft+1} && \text{(foreign NKPC)}\\
w_t &= \sigma^{-1}c_t + \nu^{-1}\ell_t && \text{(home labor supply)}\\
w_t^* &= \sigma^{-1}c_t^* + \nu^{-1}\ell_t^* && \text{(foreign labor supply)}\\
y_{Ht} &= a\ell_t && \text{(home production)}\\
y_{Ft} &= a\ell_t^* && \text{(foreign production)}
\end{aligned}
$$

Prices

$$
\begin{aligned}
p_t &= \phi_H p_{Ht} + \phi_F p_{Ft}\\
p_t^* &= \phi_H^* p_{Ht} + \phi_F^* p_{Ft}\\
\pi_t &= \phi_H \pi_{Ht} + \phi_F \pi_{Ft}\\
\pi_t^* &= \phi_H^* \pi_{Ht} + \phi_F^* \pi_{Ft}\\
\pi_{Ht} &= p_{Ht} - p_{Ht-1}\\
\pi_{Ft} &= p_{Ft} - p_{Ft-1}
\end{aligned}
$$

Resource constraints

$$y_{Ht} = \phi_H \bar{C}(c_t - \eta(p_{Ht} - p_t)) + \frac{1-n}{n}\phi_H^* \bar{C}(c_t^* - \eta(p_{Ht} - p_t^*)) + g_{Ht}$$
$$y_{Ft} = \frac{n}{1-n}\phi_F \bar{C}(c_t - \eta(p_{Ft} - p_t)) + \phi_F^* \bar{C}(c_t^* - \eta(p_{Ft} - p_t^*)) + g_{Ft}$$

Monetary policy

$$r_t^n = \rho_i r_{t-1}^n + (1-\rho_i)(\phi_\pi \pi_t^{agg} + \phi_y y_t^{agg} + \phi_g g_t^{agg}) + \varepsilon_{rt} \qquad \text{(Taylor rule)}$$

Fiscal policy

$$g_{Ht} = \rho_g g_{Ht-1} + \varepsilon_{Ht}$$
$$g_{Ft} = \rho_g g_{Ft-1} + \varepsilon_{Ft}$$
$$T_t^G = n(\bar{G}p_{Ht} + g_{Ht}) + (1-n)(\bar{G}p_{Ft} + g_{Ft}) \qquad \text{(uniform lump-sum tax)}$$

E.2 Calibration

The parameters are calibrated as in Nakamura and Steinsson (2014). For the fraction of HtM households λ, I calibrate it to match the predicted value in the census data which is 0.35. See Table E.1 for the full set of parameter values.

Table E.1: Calibration

Parameter	Target	Value
β	Discount rate	0.99
σ	Intertemporal elasticity of substitution	1
ν	Frisch elasticity of labor supply	1
χ	Preference shifter (labor supply)	1
θ	Elasticity of substitution (across variety)	7
η	Elasticity of substitution (home vs. foreign)	2
ϕ_H	Home bias	0.69
ϕ_H^*	"Home" bias of the foreign	0.03
n	Home population	0.1
a	Labor share of income	0.67
α	Calvo parameter	0.75
ρ_i	Taylor Rule: autocorrelation	0.8
ϕ_π	Taylor Rule: inflation feedback	1.5
ϕ_y	Taylor Rule: output feedback	0.5
ϕ_g	Taylor Rule: spending feedback	0
ρ_g	Government spending: autocorrelation	0.933
λ	Fraction of hand-to-mouth Households	0.35