

A New Measure of State Consumption: Construction and Applications *

Tomas Breach[†] Abhi Gupta[‡]

August 1, 2024

Abstract

The absence of a long-running measure of U.S. state-level consumption impedes the study of many questions. We address this constraint by estimating a new, annual, state-level panel of retail consumption for 1970-2015 using official consumption data and newly-digitized sales tax records. We apply our estimates to two questions whose study has been hampered by lack of data. First, we ask whether banking integration increases inter-state consumption risk sharing and find it only does so via raising output comovement. Second, using military spending shocks we estimate consumption fiscal multipliers that are positive and larger than estimates using existing consumption datasets.

*We thank Yuriy Gorodnichenko, Jon Steinsson, and David Romer for their much-appreciated guidance and Ben Schoefer, Maury Obstfeld, Emi Nakamura, Regis Barnichon, and seminar participants at UC Berkeley for their helpful comments. We also thank Jim Church of the UC Berkeley Library and Jenny Groome of the Connecticut State Library for their help. Shamal Perera, Henry Isselbacher, Paula Ore Dominguez, George Harris, Marisa Wernicke, Candice Lee, Zixun Tan, and Umar Boboyev provided excellent research assistance. This work was supported by the Clausen Center at UC Berkeley. All errors are our own.

[†]Department of Economics, University of California at Berkeley, tomas_breach@berkeley.edu.

[‡]Office of Management and Budget, abhibg94@gmail.com.

1 Introduction

What are the welfare costs of business cycles? Does government stimulus crowd out private spending? How well can households buffer shocks using credit markets? These are a few of the important questions that could be analyzed using variation across U.S. states and a good measure of state-level consumption. The literature taking a regional approach to macroeconomic questions is large (Chodorow-Reich (2020)), and much of it relies on the state versions of official national series like output, employment, and personal income. However, there is a notable gap in data availability: no long-running, official measure of state consumption exists. This absence has led researchers to work with imperfect alternatives whose shortcomings can impede the study of pressing topics like these.

To address this key issue, this paper constructs a new, annual, state-level consumption measure for all U.S. states and the District of Columbia for 1970 to 2015. Our core idea is that while existing measures of consumption might each be individually deficient, we can combine them to estimate true consumption. By choosing input data whose shortcomings are pure measurement error we can construct a consumption measure that is still broadly useful. Our final output is a state-level, nominal, per-capita retail consumption series that we benchmark against leading alternatives and official personal consumption expenditures (PCE) data. We demonstrate the usefulness of our estimated consumption series by examining two questions whose study has been hampered by poor quality data: the impact of financial integration on interstate risk sharing and the response of consumption to fiscal shocks.

The central shortcoming of existing measures of state consumption is that they require researchers to trade off data quality with coverage across time and states. While generally high quality, official government measures have limited availability. State-level PCE begins only in 1997, the economic census is comprehensive but occurs only every five years, and the state-level Monthly Retail Trade Survey (MRTS) was constructed for but a few states and has long been discontinued. In search of greater coverage, some researchers have turned to correlates of consumption such as retail employment or new car registrations. These studies often require additional structural assumptions or empirical work to assure audiences that the “noise” that separates these alternative measures from true consumption is not endogenous to

the topic of interest. Others have used composite private sector measures that are available for many years and states but are typically a “black box” combination of employment and other undisclosed inputs. The nature and extent of the measurement error they contain can therefore be hard to assess¹. There are still more alternatives depending on the setting, but, in general, compromises are hard to avoid.

With these issues in mind, we begin in section 2 by assembling and harmonizing several official measures of retail spending. This focus on retail balances coverage of the consumption basket with data quality and is a common choice in previous papers interested in state-level consumption². Our first source is the economic census, which every five years surveys the near-universe of retail establishments. It has been used infrequently in the past due to low frequency and changing definitions of retail sales over time. We use these censuses to build a novel, internally consistent series of state-level cumulative 5-year retail sales growth for 1967-2012. Our second main source is the Census’ Monthly Retail Trade Survey, which surveys a subset of retail establishments to provide a higher-frequency but noisier measure of retail sales. From 1962 to 1996, the MRTS reported retail sales estimates for a selection of sub-national areas. In the past, researchers have made use of a limited subsample that covers 19 large states and 1978-1996. We extend previous work by digitizing annual state-level MRTS data for 1970-1996. We also digitize MRTS data for the nine census divisions, giving us a regionally-aggregated measures of retail sales³. Third, we construct the retail component of state-level PCE from the goods and food services subcategories to maintain comparability with the retail census.

We also construct measures of retail spending based on state sales tax records. State sales taxes are ideally designed to target retail sales. With data on both tax rates and revenues, researchers can therefore calculate the implied tax base and use this as an estimate for retail

¹One leading example is the Survey of Buying Power from Sales and Marketing Management (SMM), which Asdrubali et al. (1996) and Del Negro (2002) have used. Guren et al. (2021) and Del Negro (2002) discuss the opacity of its construction.

²Del Negro (2002), Asdrubali et al. (1996), Guren et al. (2021), and others also focus on retail spending and discuss its connection to total consumption. In our sample national retail sales are closely correlated with PCE itself at the annual frequency and the retail component of national PCE is roughly half the total basket.

³The MRTS reported state-level estimates only for states of substantial size. The group of covered states grew to 15 by 1970 and 19 by 1978. The census divisions are clusters of neighboring states. The MRTS provides annual retail sales for all nine divisions over 1970 to 1996.

spending. This was pioneered by Garrett et al. (2005) and has been used occasionally in the literature (e.g. Zhou (2010) and Abdallah and Lastrapes (2012)). We build on their work by constructing an updated set of estimated sales tax bases for 46 states and 1970-2015 that corrects for previously mis-recorded rates and revenues. Unfortunately, the quality of these basic estimates can vary across states due to exemptions, rebates, and other idiosyncratic issues. For a subset of states with serious issues we construct a novel, improved measure of the sales tax base using newly-digitized industry-level sales tax records. Depending on the state, these give us either industry-level estimates of the sales tax base or, more notably, measures of gross sales by industry. Aggregating the industry-level data gives us our improved state-level retail spending measure.

We combine the information in our input series via a state-space model that links the true, latent level of retail consumption in each U.S. state to each of our observed series and accommodates their key features in section 3. We assume that retail consumption in each state evolves according to a simple unobserved components model along the lines of Clark (1987). Observations directly map onto latent consumption with classical measurement error. We calibrate our estimates to the retail census, giving us a consistent reference point for true consumption across our whole sample. We jointly estimate consumption for all states under the assumption that each state's consumption evolves independently, allowing for information sharing across states only via common parameters and the census-division-level data that capture retail sales for groups of states. We calculate the likelihood of our model using an extended Kalman filter, which accommodates differing levels of aggregation, measurement error, and missing observations. While our model does pose some estimation challenges due to its high dimensionality, we demonstrate that the Sequential Monte Carlo (SMC) method of Herbst and Schorfheide (2016) can succeed in this setting.

In section 4 we then evaluate our consumption measure by examining our estimates and the model's parameters directly, comparing our estimates to existing alternatives, and discussing the consequences of our estimation procedure for various applications. First, our parameter estimates suggest state consumption is the sum of a random walk and a smaller, persistent stationary component, which largely aligns with the literature on aggregate con-

sumption dynamics⁴. We show that the model is able to generate reasonable estimates of state-level consumption across states and years both in state-years with PCE or MRTS data and those that have only sales tax data. Even in situations where there is no state-year-specific information, such as states without sales taxes in the years before PCE is available, the model is able to infer consumption from temporally aggregated (the economic census) and regionally aggregated (division-level MRTS) data. Second, we compare our estimates to private sector estimates from Sales and Marketing Management (SMM), retail employment, and national retail PCE. Our estimates perform sensibly at the state level and track national PCE consistently across the full sample when aggregated. Lastly, using the simple framework of Kroencke (2017) we discuss why filtered quantities like PCE or our estimates are smoother than “raw” measures like retail employment or the MRTS. We show how this excess smoothness is larger when the input data is less informative, how this smoothness can bias results, and how this bias can be addressed by using designs like difference-in-differences that rely on the first moments of consumption.

With our consumption estimates in hand, section 5 turns to our first application: the impact of cross-state banking integration on interstate risk sharing. The interstate risk sharing literature has come to conflicting conclusions on the overall extent of risk sharing, largely driven by data quality and research design issues (Del Negro (2002)). To overcome these, we measure risk sharing via the relative synchronization of output and estimated consumption growth across state-pairs and focus our attention on the impact of banking integration on risk sharing. Our difference-in-difference design builds on recent work from Dube et al. (2023) and exploits the idiosyncratic and staggered rollout of deregulation across state-pairs⁵. We find that increasing integration increases output and consumption synchronization similarly, which suggests that consumers do not use bank credit to share risk across states. This contrasts with the substantial contribution of credit markets to interstate risk sharing identified by Asdrubali et al. (1996), who estimate this channel using the time-series properties of private sector consumption estimates.

⁴cf. Hall (1978) and the large subsequent literatures on the time-series properties of aggregate nondurable and durable consumption.

⁵Morgan et al. (2004) were the first to study the impact of interstate banking deregulation on output comovement. There is a large literature in finance using these deregulations as a source of exogenous variation (see Baker et al. (2022)).

For our second application in section 6 we estimate state-level consumption fiscal multipliers. The consumption response to government spending is a key determinant of the overall output multiplier and is an important quantity in its own right for understanding the welfare effects of government spending. We estimate multipliers over the 1970-2006 period using the state-level military spending shocks from Nakamura and Steinsson (2014). Using both the Nakamura and Steinsson (2014) short-run specification and Ramey and Zubairy (2018) dynamic specification, we find open economy relative consumption multipliers that are positive on impact and over 1 at five years out. Our results hold when using the weak-IV robust estimator from Andrews and Armstrong (2017). We contrast our conclusions to those of Dupor et al. (2023), who estimate smaller positive regional consumption multipliers using private sector consumption data from the Great Recession. Our findings are similar to those of Chen (2019), who estimates large, persistent effects of government spending on retail employment. Lastly, we consider our findings in the context of several recent models of open-economy fiscal multipliers and conclude these dynamic responses are hard to match without additional internal propagation mechanisms.

Related Literature

First, this paper is related to the small literature dealing with consumption measurement issues. Wilcox (1992) is an early and important study of the MRTS and the economic consequences of its measurement error. Garrett et al. (2005) and Zhou (2010) attempt to estimate state sales tax bases and the latter discusses discrepancies between private and governmental measures of consumption. Many papers that use alternative consumption measures such as Del Negro (2002) and Guren et al. (2021) also comment on their properties and potential shortcomings. Our work brings together the various datasets discussed in these papers and is informed by their insights. Also important is Kroencke (2017), which examines the differences between filtered and unfiltered consumption data through the lens of a simple statistical model. Our discussion of the causes, consequences, and solutions to “excess smoothness” in estimated consumption builds on his work.

Next, we connect to the literature on the estimation of macro quantities from noisy, mixed frequency, or missing data. The closest paper to ours methodologically is Schorfheide

et al. (2018), which estimates a state-space model for national consumption allowing for measurement error and mixed frequency data. While their approach is tailored towards learning about the low-frequency dynamic properties of consumption, our focus is on inferring true consumption given multiple noisy observables. Our modeling approach also has some similarities Antolin-Diaz et al. (2017) and others in the Bayesian dynamic factor model (DFM) literature that try to learn about latent states using many noisy observables of varying qualities and frequencies. We are able to estimate many more latent variables than a typical DFM because we impose sufficient structure on the mapping from state-level latent consumption to the observables. Our work also connects to the larger literature that attempts to measure GDP from noisy input sources using measurement error models. These include Aruoba et al. (2016), Pinkovskiy and Sala-i Martin (2016), and many others.

We also connect to the literature on interstate risk sharing. An early and influential paper is Asdrubali et al. (1996). Using private-sector consumption data, they estimate the extent of risk sharing between states starting in the 1960s and find a large role for credit markets in interstate smoothing. Since this credit channel is directly intermediated by banks, our finding in section 5 that banking integration does not increase risk sharing contrasts with their results. Hess and Shin (1998) and Del Negro (2002) also study interstate risk sharing using the MRTS and other alternative consumption measures, with the latter emphasizing the importance of taking into account measurement error. Further work in this literature such as Demyanyk et al. (2007) has often used income due to these issues with consumption data. Our new consumption estimates help resolve these data quality issues.

Additionally, our work is related to the large literature in finance on the causes and consequences of interstate banking deregulation, including Jayaratne and Strahan (1996), Kroszner and Strahan (1999), Morgan et al. (2004) and many others. Our empirical approach is informed by these papers and Baker et al. (2022)'s discussion of common econometric issues in this setting. In particular, our use of state-level synchronization as a measure of risk sharing is informed by Morgan et al. (2004) and Goetz and Gozzi (2022), who focus on changes in output comovement due to increased banking integration. We extend this approach to consumption synchronization using our estimates.

Lastly, we speak to the large literature on regional fiscal multipliers. As summarized in Chodorow-Reich (2019), there have been many studies considering the regional employment, income, or output responses to fiscal shocks. To the best of our knowledge, however, only Dupor et al. (2023) and Chen (2019) extend this focus to consumption. Dupor et al. (2023) estimates short-run multipliers using private-sector estimates of a subset of retail consumption and cross-sectional variation in federal spending during the Great Recession. Our estimates are estimated on the full retail spending basket and provide evidence on the size of the regional consumption multiplier across multiple business cycles. Chen (2019) extends the analysis of short-run fiscal multipliers in Nakamura and Steinsson (2014) to the dynamic setting and uses retail employment as a proxy for consumption. Our dynamic estimates using our consumption series are largely consistent with his findings and provide additional moments for models of regional fiscal multipliers to match.

2 Data

In this section we discuss the scope of our measurements and each of our major data sources. Much of the sales tax data we collect is new to the literature, but we also make important adjustments to sources used in previous work. At a high level, our goal is to reliably estimate consumer spending on retail goods, by collating information from the retail census, the monthly retail trade survey, and sales tax records. We orient our data cleaning efforts towards harmonizing these series, so that each represents variation in a similar basket of retail trade transactions.

The following subsection defines several concepts we use, and the remaining subsections discuss each major data source in detail. Table 1 offers a summary of each source, including its frequency, coverage, and qualities.

2.1 Retail Spending: A Definition

We define retail spending, or retail consumption, as sales of goods to final users plus food service transactions⁶. Our key data sources track the sales of retail trade establishments,

⁶The inclusion of food service in the definition of retail is based on the Standard Industrial Classification (SIC) system that was commonly used before 1997. Several of our key data sources, including the Monthly Retail Trade Survey, are of retail sales values that include sales of "Eating and Drinking Places" (i.e.,

which are establishments that specialize in retail transactions (e.g., grocery stores, auto dealers, apparel stores). For data sources that instead use the NAICS system, including the retail census for years later than 1997, our notion of retail includes NAICS codes in the 44-45 range, plus food services listed under NAICS 722. This final addition maintains continuity in the treatment of food services across our sample. We provide a detailed breakdown of the NAICS and SIC categories that our notion of retail sales covers in the appendix.

Although our data series are measurements of retail sales, our core applications require a measure of retail expenditures. The former measures sales made by firms in a state and latter expenditures made by residents of a state. The two will therefore differ by the size of transactions made by mail-order, e-commerce, and purchases made outside of one's home state. To mitigate these differences, we adjust our main benchmarks—the retail census changes—to exclude transactions by non-store retailers (those that specialize in mail order and e-commerce sales). We find that before 1997, the exclusion of non-store retailers makes little difference for our estimates of spending growth. As e-commerce gains popularity in the 2000s, sales by non-store retailers become an important part of sales dynamics, and we adjust for this accordingly when possible.

2.2 The Retail Census

Our first data source is the retail census, a comprehensive survey of US retail establishments on their annual sales and payrolls, which, since 1967, has been conducted every five years. For the years before 1997, we treat the retail census as an authoritative source for five-year retail spending growth rates in each state. This is because the census minimizes both sampling error (the survey is administered to the universe of establishments known to file payroll taxes) and non-sampling error related to industrial and geographic classification of individual establishments. A useful feature of the retail censuses is detailed information on sub-categories of retail trade, allowing us to construct five year growth rates for durable and non-durable sales, as well as exclude the sales of non-store retailers.

We make a new contribution to the literature by constructing internally consistent growth rates from the retail censuses. Most of the five year intervals in our sample featured a change

restaurants and bars).

Table 1: Summary of Data Sources

Series	Years Covered & Frequency	Source	Sales Covered	Regions
Retail Census	1963-2017, Every Years	Survey of Universe of Retail Trade Establishments that Report Payroll Tax	All Sales by Trade Under SIC/NAICS Definitions; Detailed 3 and 4 Digit Information	All US States over 1967-2017
Monthly Retail Trade Survey	1968-1996, Monthly	Random sample of Retail Trade Establishments	Retail Trade Sales, with some Disaggregation by Durable and Nondurable Establishments	15 Largest US States + 9 Geographic Divisions over 1968-1996; 4 More States Over 1978-1996
State PCE	1997-2021, Annual	BEA from Construction and Retail Census Wages and Salaries	Covers Retail Trade Categories, with Detail by Types of Consumption Expenditure	All US States over 1997 - Present
State Reports of Sales Tax Collections by Industry	Variable	Administrative Reports Based on Collections of Sales Tax by Stage Agencies	Retail Trade and Some Services, Typically Excluding Food for Home and Gasoline	Historical Industry Reports Exist for at least 27 States; We have fully cleaned and digitized 4
Quarterly Survey of State Tax Revenue	1966-2021, Quarterly	Census Bureau Survey of Aggregated Tax Revenues	Retail Trade, Mixed with Some Services and Intermediates and Industrial Goods	45 States, DC

in the technical definition of “retail sales.” For example, between 1967 and 1972, certain plumbing and electrical establishments were re-defined to wholesale from retail; between 1972 and 1977, certain types of financing fees were removed from the definition of sales. The “headline” figures reported in each census typically make no adjustment for these changes, and so a naive calculation of growth rates will conflate changes in economic activity with technical changes in definitions.⁷ Fortunately, in most cases, these census publications contain additional tables which retabulate the prior census using current definitions; we were typically able to construct consistent growth rates by digitizing these auxiliary tables. A full discussion of our census treatment is available in the appendix.

2.3 The Monthly Retail Trade Survey

The Monthly Retail Trade Survey (MRTS) reports regular estimates of national retail sales based on a random sample of retail trade establishments. Between April 1962 and December 1996, the survey published estimates for a selection of large geographic areas.⁸ Using tables published in the *Current Business Reports* series, we construct annual retail sales indexes for the 15 largest states of that era, along with 9 ‘geographic divisions’ that form a national partition. These series cover the years of 1967 - 1996, and for an additional 4 states we construct shorter series covering 1978 - 1996.⁹ Although the state-level MRTS series feature in previous work (e.g., Hess and Shin 98, Del Negro 02), a novel contribution of this paper is the extension the MRTS series to the 1967 - 1978 period for the 15 largest states and the of the division-level series¹⁰.

⁷It is especially important to avoid these errors for the 1997 transition from the SIC to NAICS classifications, where the discrepancy between the definitions of retail—even after food service is properly accounted for—is typically over 5% of the NAICS definition. Previous census users (e.g., Zhou used growth rates compromised by these large categorization artifacts.

⁸In private communication, Census Bureau officials had this to say about the termination of the geographic area figures: “There was additional geographic detail that accompanied the Monthly Retail Trade Survey releases up until the mid-1990s. Unfortunately, that portion of the program fell victim to budget cuts and was discontinued. That program relied heavily on direct collection of geographic data from the retailers.”

⁹See the appendix for a description of geographic divisions. See table () for the 19 states for which we have an MRTS series. In the appendix we also how—in addition to extending the series—we correct a data error in previous versions of the series.

¹⁰The Census website only hosts state-level MRTS records for 1986 onwards. Some recent work, e.g. Mian, Sufi, Verner 2019, has only used that subsample due to this limitation. It appears earlier researchers working with the MRTS data also digitized the data themselves or had an alternative digital record that is no longer available.

Relative to the retail census, the MRTS series are subject to sampling error, and contain much less detail on the subcategories of retail sales¹¹. However, they are available at a much higher frequency. Work by Wilcox (1992) suggests that our use of annual data should mitigate much of the sampling error that exists at monthly frequencies. The MRTS does directly calculate estimates of sampling variability for each year and geographic division. These are reported as coefficients of variation, or the relative root mean square error, and are typically between 2 and 8 percent of the level of retail sales. These are usually lower in states or divisions with larger populations as those regions have a greater number of surveyed establishments.

2.4 The BEA’s Personal Consumption Expenditures

The Bureau of Economic Analysis (BEA) publishes a state-level Personal Consumption Expenditures (PCE) series that begins in 1997. To maintain continuity with our other measures of retail sales growth, we use the goods and food services expenditure portion of this measure in our estimation. We call this subset “retail PCE.” For years ending in 2 and 7, state PCE draws heavily on information in the quinquennial retail trade census discussed above, but with some adjustments to more closely align sales figures with the notion of PCE used in the NIPA accounts. An advantage of using these figures over the raw retail census is that the explosion of e-commerce sales in the 21st century has made geographical sales figures a worse measure of geographical expenditures. The BEA’s PCE measure attempts to adjust for this discrepancy. For years between the retail censuses, the BEA uses geographic wage and employment data from retail trade industries to estimate fluctuations in consumer expenditures.¹² As the BEA notes, the core assumption behind this method is that changes in wages are closely linked to changes in receipts for these industries. The benchmarks in Awuku-Budu et al. (2013) suggest these wages are an informative but imperfect measure of expenditures. Given the geographical and imputation issues, we treat the PCE measure as an observation of retail sales subject to some noise.

¹¹Notably, there is typically enough disaggregation to permit us to distinguish between nondurable sales and total sales; we plan to further investigate construction of durable and nondurable retail sales estimates in future work.

¹²The data come from the Quarterly Census of Employment and Wages (QCEW) collected by the Bureau of Labor Statistics.

2.5 State Sales Tax Records

Previous work (e.g. Garrett et al. (2005) and Abdallah and Lastrapes (2012)) has used a measure of the sales tax base constructed as $\text{tax base}_t = \text{tax revenue}_t / \text{tax rate}_t$. If the sales tax base were consistently applied to all retail transactions, this would indeed be an effective record of those transactions. Unfortunately, this measure tends instead to be excessively volatile as a retail sales measure, for a variety of reasons: the tax base omits some stable expenditures (e.g., food), the sales tax base undergoes legal changes, and around 40% of the base falls on intermediate or non-consumer sales (Mikesell and Kioko (2018)). An additional problem is that the data source for sales tax revenues, the Census Bureau's *Quarterly Survey of State Tax Revenue*, was not conducted for 4 quarters in the early 1990s, producing a considerable gap for growth rate calculations.¹³. Zhou (2010) improved on these previous efforts for about 12 states by constructing a more consistent across time measure of taxable sales using state-level measures of rate changes and revenues.

To build an improved series of taxable retail sales, we collect novel data on state sales tax collections by industry. Relative to previous efforts, we focus not just on collecting reports of the tax base, but specifically reports that isolate the the receipts coming from the retail trade sector. By focusing on receipts from the retail trade sector, we can purge variation in the tax base measure that comes from non-retail transactions; in some cases the industry level data also allow us to adjust growth rates for legal changes in the scope of the tax base (e.g., the exclusion of groceries). Some of our data sources are the same ones used by Zhou, but the new reports we digitize allow us to greatly expand the time series coverage and industry detail available. For example, we extend Zhou's gross sales series for Washington State, which began in 1994, back to 1972, and are able to separate Washington's retail sales into durable and non-durable establishment series.¹⁴

To preview the benefits of our additional data, Figure (1) plots, for Washington State,

¹³The sale tax data excludes the states which do not collect sales tax: Alaska, Delaware, Montana, New Hampshire, and Oregon. Users of the *Quarterly Survey* also tend to exclude Utah and Nevada, two states with implausibly wild time series.

¹⁴The reason we are able to collect longer time series than Zhou (2010) is that she restricted her search to data available on state websites, which states only began uploading in the mid to late 1990s. Most of our new records come from an (ongoing) search of state library systems.

percent changes in the crude sales tax base (in blue), along with a sales series we construct from our auxiliary data (in red). In 1982, Washington temporarily repealed the sales tax exemption for groceries: this caused a huge surge in the crude sales tax base, followed by a large drop in the crude base when the repeal ended in 1983. Thus, the crude base conflates changes in transactions with legal changes to the transactions measured by the state. We also observe the additional problem that crude base is missing data for 1992 and 1993, when the *Quarterly Survey of State Tax Collections* was suspended for four quarters.

In contrast, our improved series allows us to correct the effects of the tax base change, by informing us directly about the revenue collected by grocery store establishments. This allows us to compute a percent change which excludes the industry whose legal status is evolving. In the case of Washington State, we can do even better than this, as the auxiliary data reports a sales measure (“gross sales”) which is invariant to changes in the tax base.

Washington: Two Measures of Retail Trade Sales, Pct Chg (1972-1997)

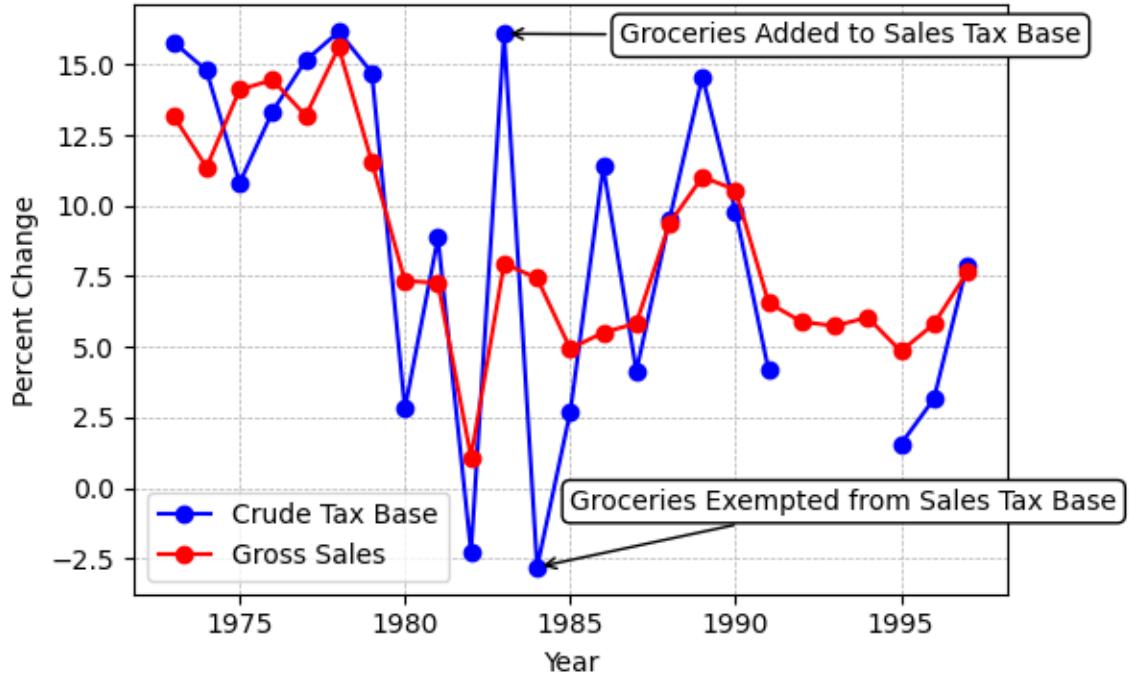


Figure 1: The blue series is a crude estimate of the sales tax base used in previous work. The red line plots our refined estimate of retail sales (based on sales tax filings), using additional taxable sales information collected from historical reports of the Washington State Department of Revenue.

In our data appendix we discuss the sources and concepts involved in our construction of improved sales tax base measures. Although we have uncovered information on tax bases by industry for over 25 states, this version of the paper is only implementing a fully improved sales tax base for 4 states, owing to the considerable effort needed to clean these series.¹⁵ Thus, for most states we are currently using a tax base series similar to the one constructed by Zhou (2010). This is the crude tax base discussed previously, although relative to previous papers we are more careful to document and exclude dates where there are legal records of a large change to the sales tax base.

3 Estimation

On their own, each of these previously discussed data series has incomplete coverage across states and years, is subject to measurement error, or is aggregated in time or across states. Our aim therefore is to estimate annual per-capita log retail consumption at the state level using the information embedded in these data series. Our output is a state-by-year panel of consumption estimates for the 50 states and D.C. for 1970-2015. We proceed by specifying a state-space model that links our observed series to the unobserved latent level of consumption in each U.S. state and estimating the model using Bayesian methods.

3.1 Model Overview

Our model has two main blocks: a latent block that governs the dynamics of consumption and an observation block that maps observed data series to latent consumption. For the latent block we model consumption using an unobserved components (UC) model. These models have long been used to decompose macroeconomic time series into stochastic trends and stationary cycles. The form used here is similar to one used in recent work by Farmer et al. (2021) to model GDP and nests the specific cases of stationarity, stationarity around a time trend, and stationary around a random walk with drift. While our setting differs in an important respect from typical use of these models by incorporating measurement error, we choose this functional form for its parsimony, flexibility, and success in modeling other

¹⁵The 4 states are Washington, Colorado, Utah, and Georgia. We choose to focus on these states, as they are relatively large but lacked information from the Monthly Retail Trade Survey.

macroeconomic time series. For the observation block, we assume classical measurement errors that are independent across series, time, and with respect to latent consumption.

Let state i 's true, latent level of log annual per-capita retail consumption c_t be the sum of a stochastic trend \bar{c}_t and cycle x_t . We have:

$$\begin{aligned} c_{i,t} &= \bar{c}_{i,t} + x_{i,t} \\ \bar{c}_{i,t} &= \bar{c}_{i,t-1} + \mu_i + \epsilon_{i,t} \\ x_{i,t} &= \rho x_{i,t-1} + \eta_{i,t} \\ \epsilon_t &\sim N(0, \lambda^{1/2}\sigma), \quad \eta_t \sim N(0, (1-\lambda)^{1/2}\sigma) \end{aligned}$$

The trend component \bar{c}_t evolves as a random walk with drift μ_i and the cycle component x_t is an AR(1) process. The variance of the increment to c_t is σ , with share λ coming from the trend shock ϵ_t and the remainder coming from the cycle shock η_t . As $\lambda \rightarrow 0$, c_t approaches a trend-stationary specification with time trend μ . As $\lambda \rightarrow 1$, c_t approaches a pure random walk with drift μ . We further assume that ϵ_t and η_t are independent.¹⁶

The observation block of the model maps latent consumption to each of our observable series with measurement error. The general structure is that each observable series j for state i has observations $y_{i,j,t}$ that are a known function of consumption and its lags $g_j(c_{i,t}, c_{i,t-1}, \dots)$ with an average deviation $\alpha_{i,j}$ and measurement error $\omega_{i,j,t}$:

$$y_{i,j,t} = g_j(c_{i,t}, c_{i,t-1}, \dots) + \alpha_{i,j} + \omega_{i,j,t}, \quad \omega_{i,j,t} \sim N(0, \sigma_j^\omega)$$

The g are allowed to be nonlinear. We allow for persistent average deviations, α_j , to capture differences in trends between observable series and latent consumption. The main source of these are the previously discussed differences between the retail census' conception of retail sales, retail sales as tracked by the MRTS, our retail PCE measure, and state sales tax bases. For example, the retail sales tax base has been eroding over time relative to total retail sales

¹⁶As Clark (1987) and Morley et al. (2003) discuss, a UC model with unrestricted covariances of fundamental shocks is not identified without auxiliary non-linear restrictions. In practice, even models where the structural covariance between trend and cycle shocks is zero can yield estimated shocks that do covary. In light of these issues, we follow previous work and impose independence.

due to accumulating exemptions, which can introduce long-run trends unrelated to true retail consumption. Lastly, we assume here that the measurement errors are classical—*independent* of the structural shocks ϵ_t and η_t —and *mutually independent*.

In particular, we observe 4 state-level series: annual changes subject to measurement error for the sales tax base and the state MRTS; annual levels subject to measurement error for retail PCE; and the 5-year changes in the retail census subject to no measurement error:

$$\begin{aligned} y_{i,SalesTax,t} &= c_{i,t} - c_{i,t-1} + \alpha_{i,SalesTax} + \omega_{i,SalesTax,t}, \quad \omega_{i,SalesTax,t} \sim N(0, \sigma_{SalesTax}^\omega) \\ y_{i,MRTS,t} &= c_{i,t} - c_{i,t-1} + \alpha_{i,MRTS} + \omega_{i,MRTS,t}, \quad \omega_{i,MRTS,t} \sim N(0, \sigma_{MRTS}^\omega) \\ y_{i,PCE,t} &= c_{i,t} + \alpha_{i,PCE} + \omega_{i,j,t}, \quad \omega_{i,PCE,t} \sim N(0, \sigma_{PCE}^\omega) \\ y_{i,RetailCensus,t} &= c_{i,t} - c_{i,t-5} \end{aligned}$$

We also observe and make use of annual census-division-level MRTS changes for each division d . Because our units are log per-capita levels, the mapping from state to division is nonlinear and depends on state populations:

$$\begin{aligned} y_{d,MRTS,t} &= g_d(\{c_{i,t}\}, \{pop_{i,t}\}) - g_d(\{c_{i,t-1}\}, \{pop_{i,t-1}\}) + \alpha_{d,MRTS} + \omega_{d,MRTS,t}, \\ \omega_{d,MRTS,t} &\sim N(0, \sigma_{MRTS}^\omega) \\ g_d(\{c_{i,t}\}, \{pop_{i,t}\}) &= \log \left(\frac{\sum_{i \in d} pop_{i,t} \cdot \exp c_{i,t}}{\sum_{i \in d} pop_{i,t}} \right) \end{aligned}$$

We take the five year changes in the retail census as the truest measure of retail consumption growth and the level of PCE as an indicator for the level of consumption.

While one could estimate 51 independent state-specific models, there are gains from sharing information across states that we seek to exploit. First, conducting inference on the parameters that govern the dynamics of latent consumption may be challenging on a state-by-state basis given our short time series¹⁷. Imposing common ρ , λ and σ across states allows us to leverage the panel dimension of our dataset effectively. Second, census-division-level MRTS data is most useful in the context of jointly estimating each constituent state's

¹⁷These issues are discussed in Farmer et al. (2021) and are potentially a greater concern here as the time dimension of our annual dataset is much shorter than their quarterly one.

latent consumption as there is no obvious way to transform such a series into state-specific information. Lastly, we expect that the measurement errors inherent in the MRTS and PCE series are broadly similar across states, making it natural to impose common σ_{PCE}^ω and σ_{MRTS}^ω .

3.2 Estimation Approach

We take a Bayesian approach to model estimation to take advantage of our prior knowledge about the properties of our data series. We start by specifying priors for the model's parameters and casting the model in state-space form. We evaluate the likelihood using an extended Kalman filter to address non-linearities in the measurement equation. We estimate the model by sampling from its posterior using Sequential Monte Carlo. We use the smoothed (i.e. two-sided) estimates of latent consumption as our output.

We compile the previously mentioned series into a single $\approx 4,000$ observation dataset spanning 51 states and 1970-2015. See table 2 for specifics and 1 for more details on the underlying data. We convert aggregate quantities into per capita ones using smoothed state-level population estimates from the Census¹⁸. Coverage is generally comprehensive. Every state-year has aggregated information available from the division-level MRTS and the retail census and for about 90% of state-years we observe state-year-specific data. While the state sales tax records are occasionally missing in our sample, the typical small state (e.g. Maine) has sales tax data, division-level MRTS, and retail census data for the whole sample and retail PCE data post 1997. Larger states (e.g. New York) will also have state-specific MRTS data. The states without sales taxes (AK, DE, NH, MT, OR) do not have state-specific MRTS data and so have no state-specific data prior to 1997. However, the division-level MRTS data does provide spatially aggregated annual information and the retail census provides 5-year cumulative changes.

The priors for our parameters are generally diffuse and informed by national statistics, reported sampling errors, or comparisons between series. A summary can be seen in table

¹⁸There are occasionally large changes in estimated population between the last inter-census estimate and the following decadal census estimates. For example, Florida grew roughly 1.5% from 1998-1999 and 2000-2001 but 6% from 1999-2000. Our smoothed population estimates are constructed to match the decadal estimates exactly.

Data Series	Units	Areas	Years Cov- ered	Observations
Retail PCE	log per capita level	51 States	1997-2015	918
MRTS: original states	log p.c. growth	CA, FL, IL, IN, MA, MI, MN, MO, NJ, NY, NC, OH, PA, TX, WI	1970-1996	390
MRTS: later states	log p.c. growth	LA, MD, NC, VA	1978-1996	76
MRTS: census divisions	log p.c. growth	9 Divisions	1970-1996	243
Sales Tax Bases	log p.c. growth	All but AK, DE, NH, MT, OR	varied (generally early 70s-2015)	1858
Retail Census	5 year log p.c. growth	51 States	Every 5 years from 1972-2012	459

Table 2: Summary of Data Used in Estimation

3. Average state-level growth rates μ_i are symmetric and centered on rounded average nominal US retail PCE growth over 1970-2015. ρ and λ have nearly-uniform Beta priors on $[0, 1]$ that drop off within 1e-3 of the bounds to rule out edge cases that can threaten identification such as $\lambda = 1$. The range restriction on ρ is to ensure stationarity of the cycle and rule out negative autocorelation¹⁹. The conditional variance of consumption growth σ is centered around the standard deviation of log national retail PCE growth with 25th and 75th percentiles of about 4 log points and 9 log points. The average measurement error parameters $\alpha_{i,j}$ are mean zero and diffuse. Priors on measurement error standard deviations are relatively diffuse and centered conservatively on larger values. Priors for σ_{PCE}^ω and σ_{MRTS}^ω have mean 0.01, i.e. one log point, to accommodate geographical differences and imputation for PCE and reported estimates of sampling errors for the MRTS. The mean for $\sigma_{SalesTax}^\omega$ is larger at .02 and is loosely based on the observed variance of the sales tax base versus the MRTS for the years and states where they overlap.

Collect the parameters of the model into $\theta = [\rho, \lambda, \sigma, \{\mu_i\}, \{\alpha_{i,j}\}, \sigma_j^\omega]$. Taking into account states that lack any MRTS or sales tax observations, and therefore have no need for $\alpha_{i,j}$ s

¹⁹Both national retail PCE and state retail PCE have strongly positive first order autocorrelations.

Parameter	Prior Dist.	25th and 75th percentile	Source/Reasoning
μ_i	$N(0.05, 0.1)$	(-0.02, 0.12)	U.S. Retail PCE growth
ρ	$\beta(1.05, 1.05)$	(0.26, 0.74)	Rules out edge cases
λ	$\beta(1.05, 1.05)$	(0.26, 0.74)	Rules out edge cases
σ	$\Gamma^{-1/2}(0.085, 0.15)$	(0.04, 0.09)	Variance of retail PCE growth
$\alpha_{i,PCE}$	$N(0, 0.05)$	(-0.03, 0.03)	Geographical differences
$\alpha_{i,MRTS}$	$N(0, 0.05)$	(-0.03, 0.03)	Basket differences
$\alpha_{i,TaxBase}$	$N(0, 0.25)$	(-0.17, 0.17)	Basket differences, base erosion
σ_{PCE}^ω	$\Gamma^{-1/2}(0.01, 0.1)$	(0.004, 0.010)	Geographical differences, imputation
σ_{MRTS}^ω	$\Gamma^{-1/2}(0.01, 0.01)$	(0.005, 0.012)	Estimates of sampling error
$\sigma_{TaxBase}^\omega$	$\Gamma^{-1/2}(0.02, 0.25)$	(0.006, 0.016)	Variance relative to MRTS

Table 3: Priors for main parameters. $\Gamma^{-1/2}$ is the root inverse Gamma distribution. $\Gamma^{-1/2}$ and N are parameterized with mean and standard deviation.

for those states and series, gives us 182 parameters in total. For each (U.S.) state we track $c_{i,t}$, five of its lags, and $x_{i,t}$. We stack all these into a single (latent) state vector s_t that has length 356 (51 U.S. states by 7 latent states), evolves linearly, and maps onto stacked period-t observations y_t non-linearly according to the equations in the previous part:

$$s_{t+1} = T(\theta)s_t + C + R(\theta)\varepsilon_t, \quad \varepsilon_t \sim N(0, Q(\theta))$$

$$y_t = g(s_t, \theta) + D(\theta) + \omega_t, \quad \omega_t \sim N(0, E(\theta))$$

For more information on the structure of these matrices see the appendix. We calculate likelihoods by casting the model in state space form and applying an extended Kalman filter, which adapts the standard Kalman filter straightforwardly to nonlinear settings using a period-by-period first-order approximation of g . Here, $g(s_t, \theta)$ is linear except for the population-weighted transformation that maps state-level latent consumptions to division-level MRTS growth rates²⁰. The extended Kalman filter handles missing observations by subsetting out the corresponding rows in $g(s_t, \theta)$, $D(\theta)$ and ω_t as needed period-by-period.

We estimate the model using Sequential Monte Carlo, an alternative to Random Walk Metropolis Hastings that has previously been applied to estimating medium-scale DSGE

²⁰Nonlinearity arises from the time-varying population weights, which we take as given, and the log-averaging transformation itself. In our testing, the first-order approximation is quite accurate.

models. The implementation we use here was developed by Herbst and Schorfheide (2016) and has been applied by Cai et al. (2021) and others. SMC can be viewed as an augmented form of importance sampling in which draws from a target distribution, here the model’s posterior, are created by reweighting draws from an easier to sample distribution, here the model’s prior. SMC turns modifies this one-step approach by incrementally and adaptively mutating a set of initial parameter draws from the prior into draws from the posterior. Recent work in statistics on related particle filtering methods has proven that the adaptive features of Herbst and Schorfheide’s SMC implementation ensure performance in high dimensional settings (Rebeschini and Van Handel (2015)).

The output of SMC is a set of N parameter draws and weights $\{\theta^i, W^i\}, i = 1\dots N$ that approximates the posterior. We choose N in conjunction with other tuning parameters to balance sample quality and execution time following the discussions in Herbst and Schorfheide (2016) and Cai et al. (2021). Our nominal sample size is $N = 1000$ and our effective sample size per parameter is in the low hundreds. This compares favorably to standard random walk Metropolis Hastings estimations of medium-scale DSGE models where nominal sample sizes in the millions can translate to fewer than 50 effective draws²¹. We provide further estimation details in the appendix.

4 Evaluation of Consumption Estimates

First, we relate our posterior parameter estimates to existing research on consumption and our prior knowledge about the input series. Second, we compare estimated consumption across states and years with differing levels of information. We then compare our estimates to retail employment, the SMM estimates, and national PCE. Lastly we describe the general properties of filtered estimates versus “raw” ones and consider the implications.

²¹Herbst and Schorfheide (2014) presents several examples to this point. The numerical error inherent in a Monte Carlo average is inversely proportional to the effective sample size and varies by parameter. Intuitively, parameters that are more “difficult” to learn about from the data have smaller effective sample sizes and more numerical uncertainty.

4.1 Posteriors for Parameters

Before examining the posterior estimates directly, it is helpful to take a high-level view of how these parameters are intuitively identified from the data we have. First, PCE and the retail census pin down the level and average growth rates μ_i of $c_{i,t}$, respectively. The census is especially important as it has complete coverage across states and 5 year periods. Second, our priors and the overlap between various series inform the measurement error variances σ_j^ω and average deviations $\alpha_{i,j}$. For example, the overlap between sales tax bases and PCE in most states and sales tax bases and the retail census over time inform $\sigma_{SalesTax}^\omega$. Lastly, the parameters ρ , λ and σ are estimated from the dynamics of estimated consumption over time. As discussed in Farmer et al. (2021), these parameters are often quite difficult to learn about in typical macro sample sizes. Since we impose that they are common across states we leverage the panel structure of our data to effectively increase the number of observations.

We report posterior estimates for the model’s core parameters in table 4. We consider two main subsets of the parameters. First, the parameters governing the dynamic behavior of latent consumption— ρ , λ and σ —conform with existing work on consumption dynamics. We estimate that the level of state-level consumption is very close to a random walk, with a small and persistent AR component whose contribution to overall variation ($1-\lambda$) is imprecisely estimated at around 10-30 percent. We can also see this by looking at consumption growth, $\Delta c_t = (\rho - 1)x_{t-1} + \mu + \epsilon_t + \eta_t$, which is dominated by time t shocks and has little persistence given our estimate for ρ is about 0.965. There is a large literature on the dynamics of aggregate consumption started in its modern form by Hall (1978). Given our frequency is annual and we include both durables and nondurables in our consumption basket, our estimated mixture of a strong random walk and a smaller persistent mean-reverting component fits within the existing literature²².

Second, our posterior means for the measurement error shocks are consistent with previously discussed properties of our observable series including sampling errors, geographical

²²As discussed in Kroencke (2017), there is ongoing disagreement in the asset pricing literature as to the extent and importance of a stationary component in nondurables consumption. The literature on the aggregate dynamics of durables spending has also noted that it has a persistent component (e.g. Caballero, 1993).

and imputation-based differences, and general measurement error. The MRTS is the closest in construction and coverage to the retail census and as such it is not surprising that we estimate low levels of measurement error for it. The posterior mean for σ_{MRTS}^ω is near the lower end of reported sampling error for the series. The level of σ_{PCE}^ω is higher but is consistent with the series' observed properties. For the states and years where they overlap, we calculate that the difference between the 5-year growth in retail PCE and corresponding the retail census observation has a standard deviation of about 4 log points. While not directly mapping to σ_{PCE}^ω , this moment suggest that the previously-discussed differences between PCE and the retail census are quantitatively large. Lastly, the standard deviation of sales tax base growth is about 1 log point. Our sales tax measurement error $\sigma_{SalesTax}^\omega$ implies that it would not be unusual for the sales tax base to grow by 5 log points while true consumption grew by 3.

Parameter	Description	Mean	25th/75th
ρ	Persistence of transitory component of $c_{i,t}$	0.964	0.963, 0.967
λ	Variance share of permanent shocks to $c_{i,t}$	0.870	0.65, 0.92
$100 \cdot \sigma$	Std. dev. of total shocks to $c_{i,t}$	6.8	6.65, 6.91
$100 \cdot \sigma_{PCE}^\omega$	Meas. error for retail PCE	0.33	0.31, 0.34
$100 \cdot \sigma_{MRTS}^\omega$	Meas. error for the MRTS	0.07	0.06, 0.08
$100 \cdot \sigma_{TaxBase}^\omega$	Meas. error for the sales tax base	0.97	0.96, 0.99

Table 4: Posterior estimates for main parameters

Our estimates of consumption are the smoothed, i.e. two-sided, values of latent consumption. Formally, we use as our estimates $E[c_{i,1:T}|y_{1:T}]$ where $y_{1:T}$ is the full set of data used in estimation. We calculate this expectation by adapting the simulation smoother of Durbin and Koopman (2002) to our non-linear setting and taking the weighted average of smoothed latent consumption across parameter draws²³. Details of the smoothing procedure, plots of the distributions of parameters, and plots showing the uncertainty in our estimates are in the appendix.

²³We adapt the original Durbin-Koopman smoother similarly to how the extended Kalman filter adapts the standard Kalman filter. The weights are the W^i calculated by SMC.

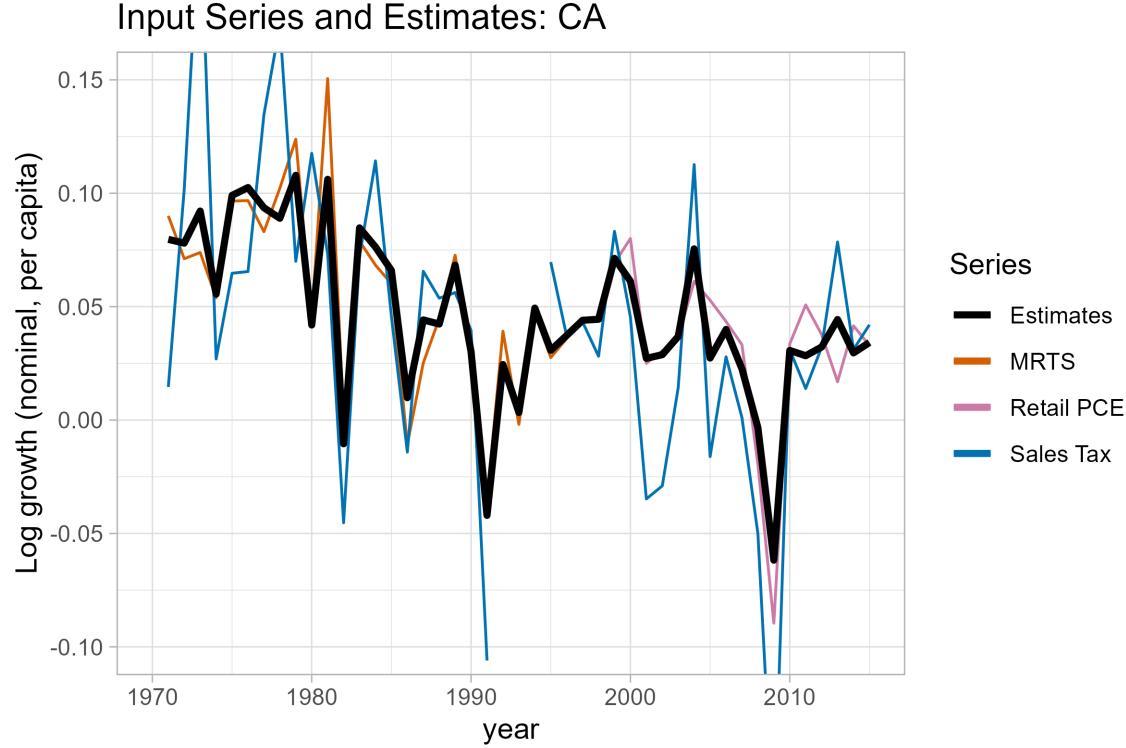


Figure 2: Comparison of estimates and input series for California. All states make use of division-level MRTS (unplotted), the 5-year retail census (unplotted) and state retail PCE. California’s estimates also use a state-level MRTS series and a basic sales tax base series constructed from tax rates and revenues.

4.2 Overview of Estimates and the Impact of Data Availability

To get a sense of how the model performs, it is helpful to first consider estimates from a state that has all five input series available. In figure 2 we show the input series and estimated latent consumption for California. It has state-level MRTS, sales tax base, and PCE observations in addition to the unplotted division-level MRTS and retail census. The sales tax measure for California is a basic version based on reported rates and revenues and has occasional missing periods. Following the parameter estimates, we see that estimated consumption hews more closely to the MRTS than the sales tax base in the early part of the sample. In the latter part of the sample estimated consumption does tends to track PCE closely, putting little weight on particularly large moves in the tax base as seen in the early 2000s. California’s experience is typical for the 19 states that have MRTS observations—even in periods where the sales tax base behaves oddly, the low measurement errors assigned

Input Series and Estimates: CO

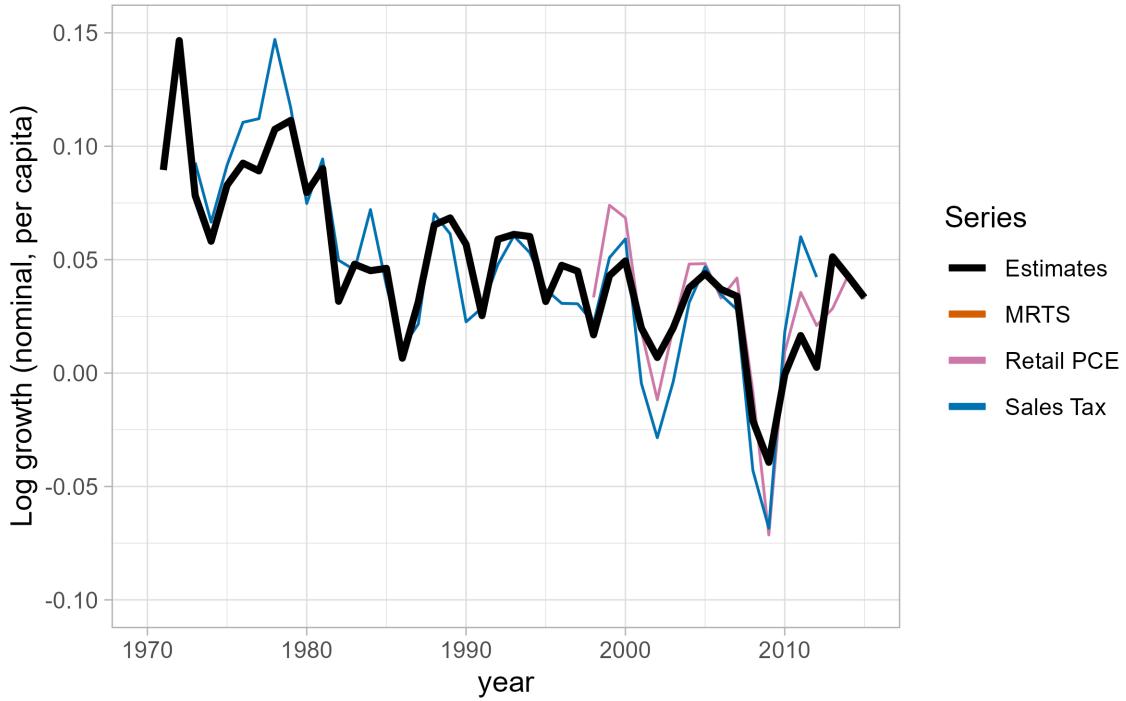


Figure 3: Comparison of estimates and input series for Colorado. All states make use of division-level MRTS (unplotted), the 5-year retail census (unplotted) and state retail PCE. Colorado also has an improved sales tax base series constructed from detailed state records.

to the MRTS and PCE keep the estimates close to these series.

Next, consider states that lack the MRTS but have sales tax data, such as Colorado in figure 3 or Rhode Island in figure 4. Colorado uses an improved sales tax series constructed using industry-level gross sales instead of cruder sales tax rates and revenues. Estimated consumption remains close to PCE in the latter part of the sample and close to the sales tax measure in the earlier half. It is identical to neither, however, because division-level MRTS and retail census information is also incorporated. For states with a cruder measure of the state sales tax base, which are typically more volatile, the estimates lean more on the division MRTS and retail census. For example, Rhode Island, shown in figure 4, has a volatile and cyclic sales tax base from 1970 through the late 80s. The model interprets much of this variation as measurement error-driven, yielding estimates of consumption that are smoother than the sales tax series. Simultaneously, the model also adjusts the longer run trend of consumption growth to match the retail census instead of the trend present in

Input Series and Estimates: RI

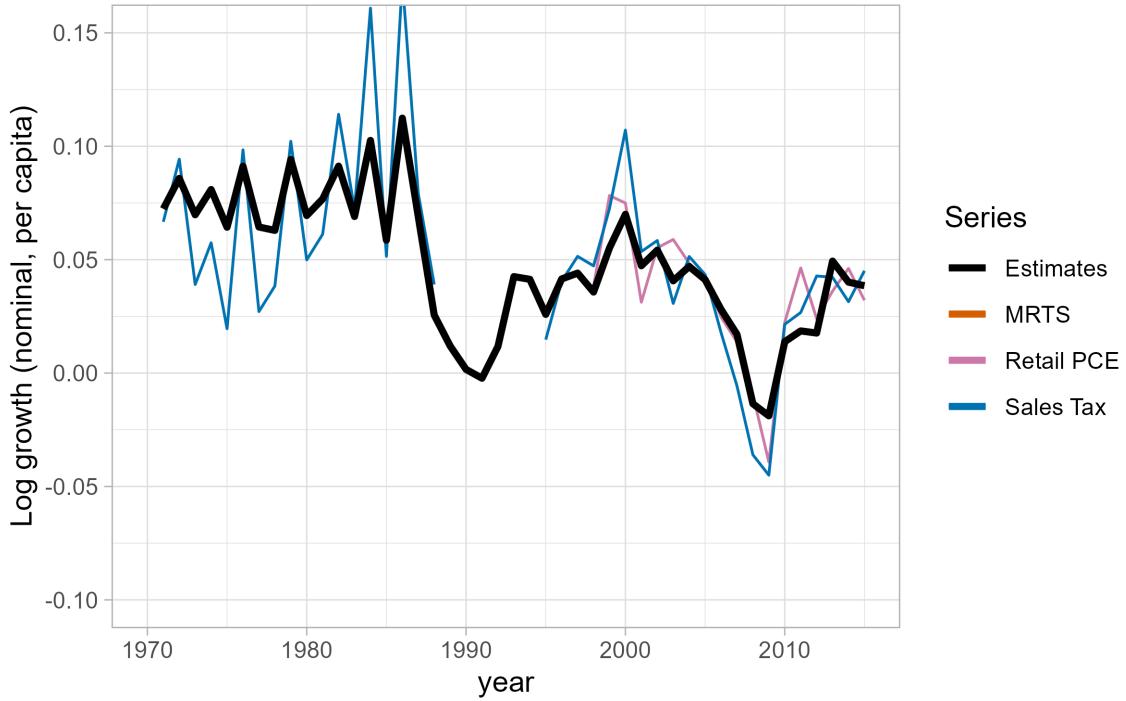


Figure 4: Comparison of estimates and input series for Rhode Island. All states make use of division-level MRTS (unplotted), the 5-year retail census (unplotted) and state retail PCE. Rhode Island has a standard sales tax base series constructed from tax rates and revenues.

sales taxes in the 70s and 80s.

Lastly, consider state-years where we have no state-specific information. In our dataset, about 90% of state-years have state-specific data, with about half of the missing state-years coming from states without sales taxes and the remainder from states idiosyncratically missing sales tax data. For example, there is a multi year period in the late 80s to early 90s for which Rhode Island has no state-specific data due to missing sales tax revenue information. Despite this, the model infers non-trivial state-level dynamics using temporally and regionally aggregated information from the retail census and division-level MRTS, respectively. The five states that are missing sales tax information completely²⁴ do have non-trivial longer-run consumption growth dynamics but generally quite smooth consumption growth year-to-year. In sum, the model is capable of reconciling conflicting information from across series, accommodating the difference in apparent measurement error between state sales tax bases

²⁴These states are Alaska, Delaware, Montana, New Hampshire and Oregon.

and PCE or the MRTS, and inferring state-level dynamics using temporally or regionally aggregated information when no state-specific information is available.

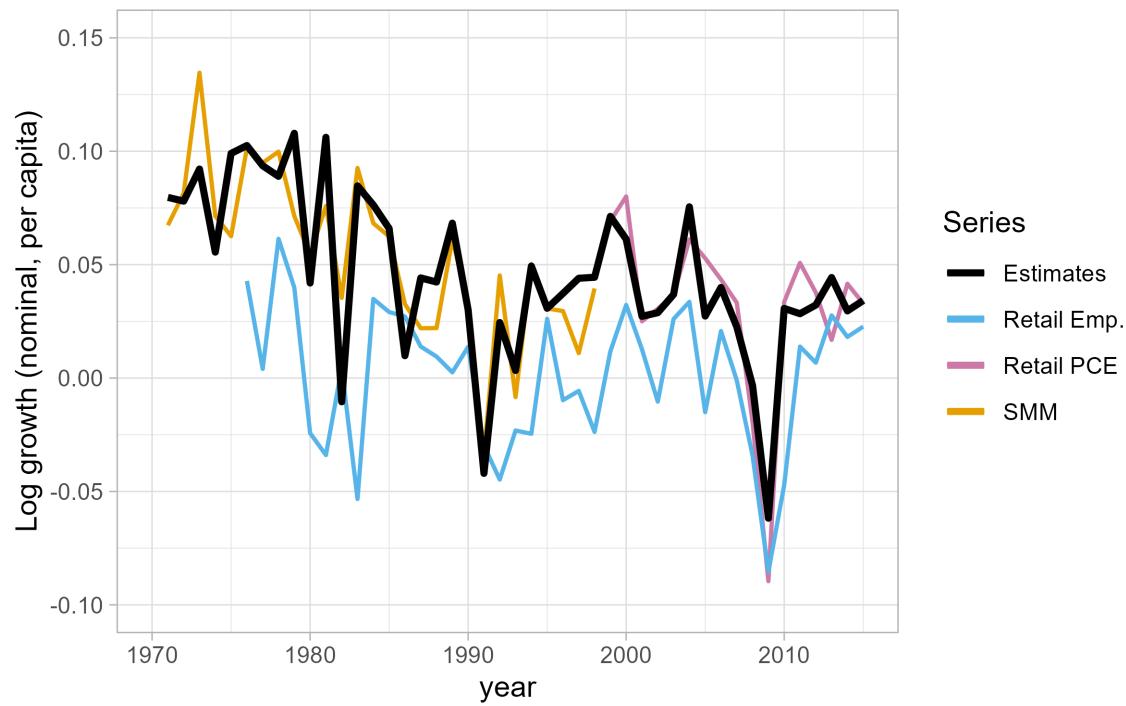
4.3 Comparison with Alternative Measures

We first compare our estimates at the state level to two leading alternative measures of consumption. First, we consider retail employment per capita. This is a commonly used measure of consumption as retail employees are an essential intermediate input into production, i.e. sales, at retail establishments. As discussed in Guren et al. (21), retail employment tracks aggregate PCE well at the national level and moves nearly one-for-one with consumption at the city level. Our annual state-level measure of retail employment comes from the adjusted County Business Patterns dataset of Eckert et al. (2021), who create a consistent NAICS-based measure of employment by sector for each county starting in 1976. It is important to note that CBP data captures employment as of March in each year, not an annual average. We aggregate these county numbers for sector 44-45 (retail trade) to create state-level retail employment and normalize by our population estimates.

Second, we consider the Survey of Buying Power from Sales and Marketing Management (SMM). These are private sector estimates of aggregate state-level retail consumption for all states starting in the 1930s and ending in 1998. Unfortunately, it is not clear how these are constructed. In previous correspondence with the firm, Guren et al. (21) and Zhou were both told that retail employment was an important, but not sole, input into the estimates. These authors were able to learn little else about their construction. Del Negro (98) compares this series to the MRTS and judges it to have roughly similar measurement error. We normalize these by our population estimates as well.

Our estimates comove moderately with these alternatives on average with some heterogeneity by state and year. In the appendix we show the panel-adjusted autocorrelation functions for our estimates of consumption, SMM's estimates, and retail employment. The within-period correlations of estimated consumption growth with SMM growth and retail employment growth are both roughly 0.5. This varies by state; in California SMM and our estimates move quite similarly, while in Colorado there are periods where they diverge substantially. In California this concordance likely arises from the California-specific MRTS

Estimates and Comparison Series: CA



Estimates and Comparison Series: CO

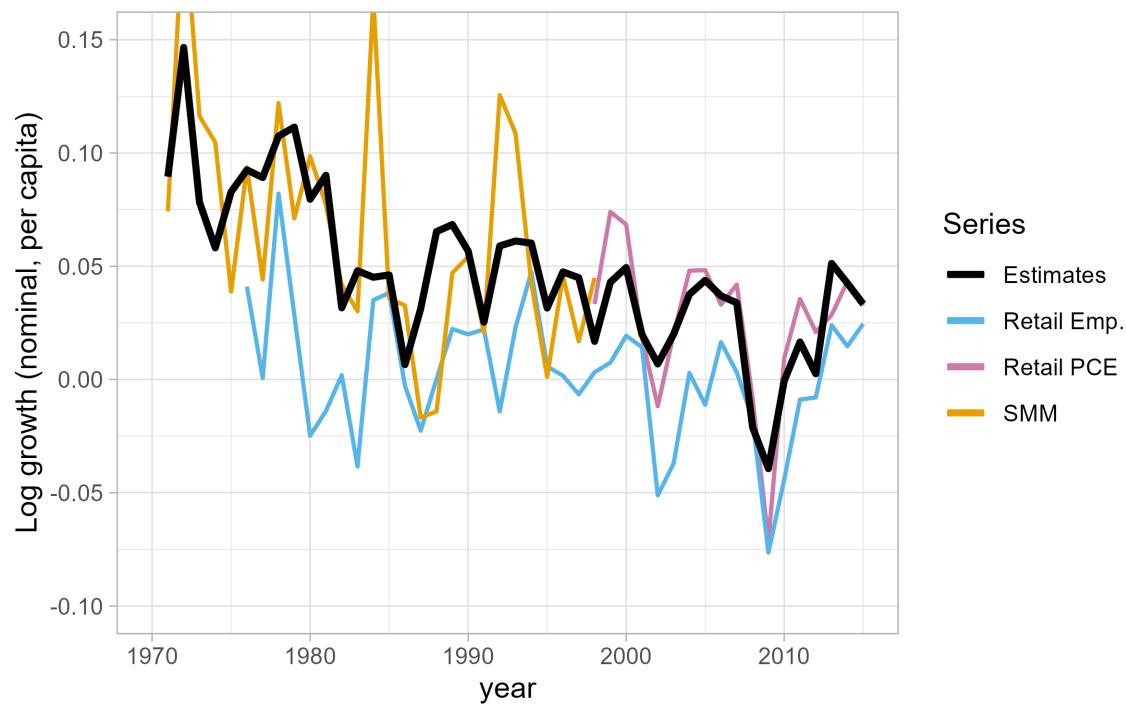


Figure 5: Comparison of our estimates, the Sales and Marketing Management series, retail employment per capita, and retail PCE for California and Colorado.

Estimates and Comparison Series: RI

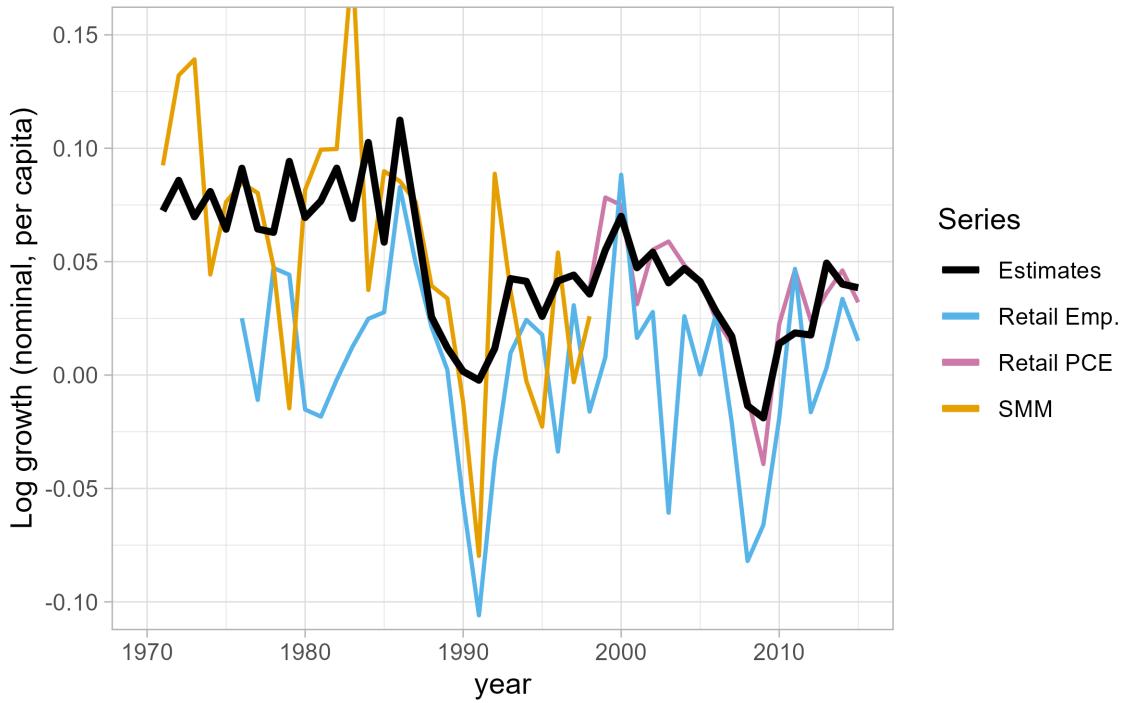


Figure 6: Comparison of our estimates, the Sales and Marketing Management series, retail employment per capita, and retail PCE for Rhode Island.

data as both our estimates and SMM's closely track the series. These are shown in figure 5. In situations where we have less or poorer quality input data, such as in Rhode Island or Delaware as shown in 6, our estimates tend to be less volatile than either alternative. Finally, our series tends to move less in years where the alternatives move substantially, such as in the Great Recession²⁵.

We also compare our estimates and these two alternatives to national retail PCE. We construct the national per capita equivalents of the state-level estimated consumption, retail employment, and SMM series. The correlation of our estimates and national retail PCE is about 0.8, higher than the 0.65 for aggregated SMM and 0.6 for retail employment. Our estimates also track national PCE similarly in the years when state PCE is available and when it is not with a correlation of 0.86 after PCE becomes available in 1997 and 0.78 before.

²⁵Another notable point of divergence is in Rhode Island in 1991. We estimate a very shallow decline in consumption while employment and SMM fall substantially. One potential reason for this is that the trough of the 1990-1991 recessions occurs in 91Q1 and the CBP employment data is collected every March. This is also connected to the filtering discussion in the next subsection

We take this as a sign that the combination of state sales tax records, the MRTS, and the economic census can reliably recover information about consumption. We do not track retail PCE exactly in the post 1997 period because the model still assigns some measurement error to state PCE and therefore also incorporates information from the economic census and sales tax records to generate estimates post-1997. Recall that the measurement error in PCE primarily comes from differences in retail sales within a state and consumption expenditures by state residents. We include plots of the national series in the appendix.

4.4 Excess Smoothness in Estimated Consumption

The basic theory of Kalman filters and smoothers (see, e.g., Hamilton (1994)) establishes that our estimates $\hat{c}_{i,t}$ are unbiased and minimize the mean squared error taking the data and model as given. We can also straightforwardly construct optimal estimates of linear functions of *first moments* of consumption using these $\hat{c}_{i,t}$ s. For example, growth rates can be estimated as $\widehat{\Delta c}_{i,t} = E[c_{i,t} - c_{i,t-1}|y_{1:T}] = \Delta \hat{c}_{i,t}$. In general, however, optimal guesses of *second moments* of consumption or functions thereof cannot be computed similarly; for example, given some variable x_t , $Cov(\hat{c}_{i,t}, x_t)$ is not an optimal guess for $Cov(c_{i,t}, x_t)$ nor is $V[\Delta \hat{c}_{i,t}]$ one for $V[\Delta c_{i,t}]$. Since many research questions concern second moments of consumption, this is can be problematic.

As discussed in Kroencke (2017), second moments of $\hat{c}_{i,t}$ are biased because any output of a filtering procedure will necessarily be *smoother* than the underlying true series. Recall that the Kalman filter's estimates each period are a weighted average of the incoming data and the model's prediction based on past data. The weight on the new data is the *Kalman gain* and depends on the relative variances of fundamental shocks to consumption and measurement errors. Intuitively, the Kalman gain is larger if the data is less noisy and therefore more informative about true consumption. Conversely, noisier data will lead to a smaller Kalman gain and smoother estimates that rely more strongly on the model's internal dynamics. As long as measurement error is nonzero, this excess smoothness will be present to some degree.

Furthermore, estimates are smoother when data is missing. In the appendix we show this formally in a simplified version of our model that features a single latent consumption variable and multiple noisy observables. The basic idea is that a missing observation is

equivalent to one where the measurement error is infinite: both give the model no useful information about true consumption. This can be seen visually as well in the estimates for Rhode Island in figure 4. Between 1988 and 1995, when no state-specific information is available, consumption estimates are much smoother than they are otherwise. In general, states-years with fewer observed data points will be smoother than state-years with more data.

This excess smoothness in estimated consumption can introduce bias in applications of the data. To fix ideas, consider a standard risk sharing regression along the lines of Obstfeld (1994) or Asdrubali et al. (1996) where filtered consumption is regressed on output y_t : $\hat{c}_t = \alpha + \beta y_t + \epsilon_{i,t}$. The researcher running this regression has the estimand $\beta_{\text{true}} = \text{Cov}(c_t, y_t)/V(y_t)$ in mind. However, if she uses *estimated* consumption she will instead recover:

$$\begin{aligned}\beta_{\text{filtered}} &= \text{Cov}(\hat{c}_t, y_t)/V(y_t) \\ &= \underbrace{K \frac{\text{Cov}(c_t, y_t)}{\text{Var}(y_t)}}_{\text{attenuated true coefficient}} + \underbrace{\sum_{j=1}^{\infty} (1-K)^j K \frac{\text{Cov}(c_{t-j}, y_t)}{\text{Var}(y_t)}}_{\text{bias from higher order covariances}}\end{aligned}$$

where $K \in (0, 1)$ is the Kalman gain. We derive this expression and give further discussion in the appendix. The bias introduced by using consumption estimates comes from the presence of measurement error in each period's observations and the model's use of current and past data to infer c_t . The magnitude of this bias is larger when the gain K is smaller. In practice, since the informativeness of the data can vary across states and years, the extent of the issue will vary as well. Applications where estimated consumption is used as a regressor will also be biased for similar reasons.

A variety of research design choices can minimize the impact of this excess smoothness. First, using estimated consumption as a regressor instead of a regressand will introduce further attenuation bias on top of the issues discussed above. If estimated consumption must be a regressor, then standard error-in-variables solutions such as finding an instrumental variable may be employed. Second, some regression designs that use consumption on the “left-

hand side” have estimands that are functions of first moments rather than second moments. While the estimand from the risk sharing example is a ratio of second moments, the estimand of a standard difference-in-difference design naturally takes the form of differences in means of the dependent variable. Because these means are still first moments of consumption, they are unaffected by excess smoothness concerns. We revisit this point in section 5. Third, controlling for lags of c_t will “knock out” the corresponding terms in the bias sum²⁶. Lastly, researchers can limit the sample to states and years where the data is sufficiently informative for the desired application. For example, it may be desirable to drop the states without sales taxes (DE, NH, OR, AK, MT) if the question at hand is about higher-frequency changes in or second moments of consumption.

5 Risk Sharing

In this section we estimate the impact of banking integration on risk sharing across U.S. states. First, we cover how the deregulation of interstate banking increased banking integration across state lines. We next outline how banking integration can increase risk sharing via the *credit channel*, in which an increase in credit supply helps firms and household smooth local shocks. We lastly present evidence from a difference-in-difference design and conclude that our results provide little evidence that banking integration directly increases consumption risk sharing.

5.1 Banking Deregulation and Integration

From the 1950s through the late 1970s, interstate banking was severely curtailed in the United States. The Bank Holding Company Act of 1956 required explicit state authorization for any bank seeking acquire a new out of state subsidiary²⁷ At the time, no state exercised this option and so interstate banking was prohibited except for a few, small legacy bank holding companies.

Interstate banking restrictions were progressively rolled back at the level of state-pairs

²⁶These terms change to $Cov(\nu_{t-j}, y_t)$ where $\nu_t = c_t - \hat{c}_t$ are the estimation errors. These errors are likely to be uncorrelated with y_t . We discuss this further in the appendix.

²⁷Under the McFadden Act of 1927 state banking laws restricting interstate banking applied to both state and federally charted banks.

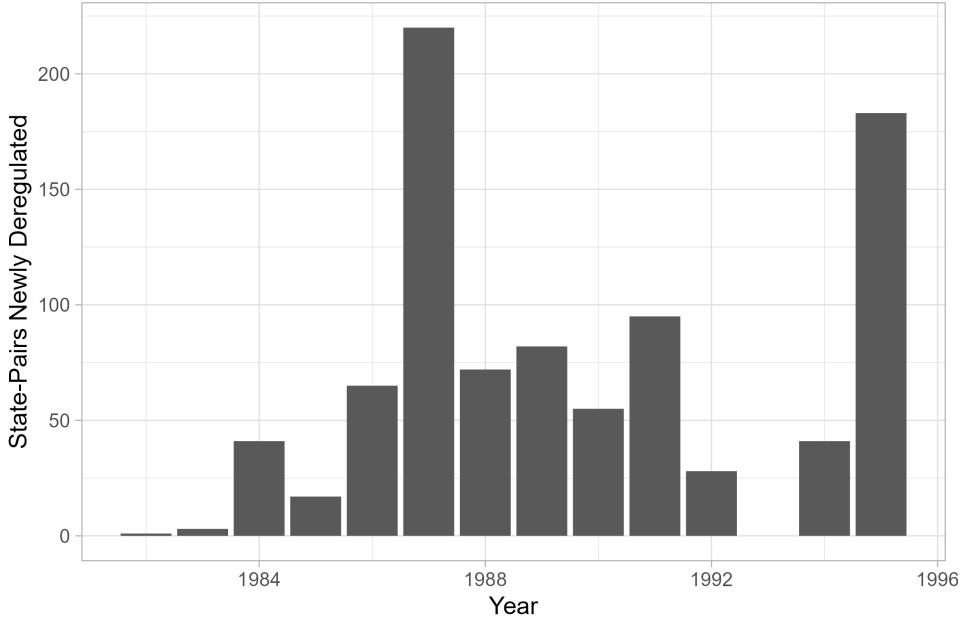


Figure 7: Number of state-pairs that deregulated interstate banking each year. After 1994, federal legislation deregulated interstate banking nationwide. We record these state-pairs as having a deregulation date of 1995.

in the 1980s and 90s. The first mover was Maine in 1978, which allowed out-of-state banks to acquire Maine banks only under the promise of reciprocity²⁸. No state actually did reciprocate until 1982, when New York and Alaska passed similar laws. Many states would follow in the coming years. Not all states demanded reciprocity; by 1984 the majority of new state-pair deregulations are coming from unilateral openings that allowed entry from without reciprocation (Goetz and Gozzi (2022), Amel (2000)). A typical progression is that a given state opens up to a select few partners in the initial round of deregulation and then progressively expands the list of allowed states and relaxes demands for reciprocity over time (Mian et al. (2020)). Amel (2000) fully documents the rollout of these laws from 1978 through 1994, when the federal Riegle-Neal Interstate Banking and Branching Efficiency Act deregulated interstate banking everywhere. Figure 7 displays the number of pairs that deregulated each year. State-pairs that were deregulated by federal legislation are recorded as having a deregulation date of 1995.

Importantly, existing research has shown that the timing of these deregulations were not

²⁸That is, a bank from state i could acquire a Maine bank if and only if a Maine bank could acquire one from state i .

endogenous to local conditions. Kroszner and Strahan (1999) and Kroszner and Strahan (2014) study the political dynamics underlying these deregulations and find a minimal role for “public interest” concerns such as a desire to grow the economy or stabilize the local banking system. The primary political opponents of deregulation were small banks and small-bank-dependent firms. Initially, these interest groups were powerful enough to defend their legal protections in every state despite wide cross-state variation in their prevalence. As the 1970s progressed, however, the nationwide development of ATMs and general improvements in information technology lowered the value of having a local bank to depositors and thus the market power of small banks. In states where these banks were initially less important the balance of power shifted earlier and interstate banking was deregulated sooner. Subsequent research from Goetz and Gozzi (2022) and others has reaffirmed that the timing of deregulation appears largely unrelated to past or expected output growth, personal income, or local banking crises.

Quantitative evidence also suggests that banking deregulations meaningfully raise the level of interstate banking integration. The degree of integration can be measured by the interstate asset ratio for states i and j :

$$int_{i,j,t} = \frac{\text{assets held by banks spanning } i \text{ and } j}{\text{bank assets in } i + \text{bank assets in } j}$$

This ratio measures the fraction of total bank assets in states i and j possessed by bank holding companies that operate in both states. It can be constructed using bank call reports, which starting in 1976 give branch-level assets and ownership for U.S. banks. This measure has been used by Morgan et al. (2004), Mian et al. (2020), and others; we use the version recently constructed by Goetz and Gozzi (2022). These authors have shown under various empirical approaches that integration is stable pre-deregulation and then rises afterwards as banks gradually expand across state lines. In table 5 we present simple evidence of this effect. We report the average levels of this ratio across “control” state pairs that were never deregulated prior to the Riegle-Neal Act and for “treatment” state pairs both before and after deregulation for the 1976-1994 period²⁹. These basic means show that integration is

²⁹The information needed to construct the integration measure is only available in call reports starting in 1976.

higher post deregulation; we estimate this effect formally in the appendix.

5.2 Banking Integration and Risk Sharing via Credit Markets

Theoretical work from Morgan et al. (2004) suggests that cross-state banking integration might increase or decrease credit local availability and, in turn, risk sharing. In their model, which adapts Holmstrom and Tirole (1997)'s framework to a multi-region setting, the equilibrium quantity of credit is determined by the banking system's supply of loanable funds and local business' and households' demand for loans. Pre-deregulation, local supply and demand shocks to the loan market do not propagate across state lines. Once interstate banking is permitted banks can use their internal capital markets to move funds between states in search of higher returns, enabling either the dampening or amplification of local shocks. The former occurs in response to local shocks to the supply of funds, which banks can temper by importing funds from elsewhere; the latter occurs in response to local shocks to loan demand, which lower local returns and push locally supplied loanable funds to different markets. Which effect dominates is an empirical question.

In the U.S. context, empirical evidence shows that integration increased credit availability to both firms and households. Morgan et al. (2004) and Demyanyk et al. (2007) provide evidence from indirect measures that lending to firms expanded in the wake of deregulation. Their results suggest that integration dampens the impact of local shocks on credit provision to firms. Mian et al. (2020) similarly find that integration increased credit availability to households by measuring mortgage and consumer loan volumes³⁰.

While an increase of household credit could increase risk sharing, the simultaneous increase in business credit can make it hard to identify the true effect. At its core, risk sharing is the idea that households' *consumption* can be insured against local, idiosyncratic shocks to *output* through linkages to other regions. Banking integration directly impacts consumption via what Asdrubali et al. (1996) term the credit channel of risk sharing: bank-intermediated household saving and borrowing across state borders can stabilize their consumption. However, banking integration also stabilizes output via a similar expansion of lending to busi-

³⁰These results do not hold in all contexts. For example, Kalemli-Ozcan et al. (2013) finds that increased banking integration across European countries in the early 2000s actually decreased credit provision to firms.

nesses, as has been shown in Demyanyk et al. (2007) and Morgan et al. (2004). To the extent that consumption is chosen *in response* to output realizations, changes in consumption dynamics conflate both the direct effect of credit expansion on risk sharing and the indirect effect operating through output. Alternative solutions to this issue include those of Del Negro (2002) and Morgan et al. (2004), who sidestep these issues by imposing much more structure and by focusing solely on output, respectively. In this paper, we address this by separately estimating the impact of banking integration on consumption and output in order to identify both the direct and indirect effects of banking integration.

5.3 Empirical Design

We use an difference-in-difference approach to estimate the impact of increasing financial integration on risk sharing. Our sample is annual, spans 1976-1994, and covers 903 state pairs. Due to excess smoothness concerns, we exclude from our analysis the five states without sales taxes (NH, DE, MT, OR, AK)³¹. Following the previous literature on interstate banking deregulation, we also drop HI, DC, and SD, due to their remoteness, small size, and tax-haven status respectively³².

Our dependent variables are output and consumption growth synchronization across state-pairs. We measure consumption using our estimates and output using nominal gross state product. Synchronization for a variable x between states i and j is defined as:

$$synch_{i,j,t}^x = -|\Delta x_{i,t} - \Delta x_{j,t}|$$

Synchronization is bounded from above by 0 and is smaller (more negative) the larger the difference in growth rates between the two regions. While alternative measures such as rolling correlations or instantaneous quasi-correlations require estimates of variances or covariances, synchronization requires only within-period estimates of growth rates. Since these first moments are estimated without bias and we use synchronization solely as a regressand, the

³¹Including these states gives point estimates that are qualitatively similar but generally attenuated and more often insignificant.

³²DE and SD have a disproportionate number of bank holding company headquarters due to their favorable tax systems. This makes it difficult to identify the portion of assets registered in these states connected to actual local banking activity.

impacts of excess smoothness in our estimates are mitigated.³³.

Table 5 gives summary statistics for the (unstandardized) banking integration measure and synchronizations of output and estimated consumption. We plot the average level of synchronization over time for each of these series in the appendix. We refer readers to Goetz and Gozzi (2022), Kroszner and Strahan (2014), and Morgan et al. (2004) for further discussion of deregulation and banking integration.

Group Period Variables	Control		Treated		Whole Sample
			Before	After	
	mean (s.e.)	mean (s.e.)	mean (s.e.)	mean (s.e.)	
Banking Integration	0.0062 (0.035)	0.011 (0.047)	0.055 (0.11)	0.022 (0.07)	
Output Synch.	-2.3 (2.1)	-3.3 (3.3)	-2.4 (2.2)	-2.8 (2.8)	
Cons. Synch.	-2.8 (2.3)	-3.3 (2.9)	-3 (2.7)	-3.1 (2.7)	
Observations	3,477	8,889	4,791	17,157	

Table 5: Summary statistics on banking integration and synchronization measures for 1976-1994. The control state-pairs are those for which interstate banking was banned until 1995 and the treated state-pairs are those that were deregulated before 1995.

We base our approach on Dube et al. (2023), who develop a straightforward local-projections based methodology that transparently addresses the major issues recently identified with standard two-way-fixed effects (TWFE) regressions³⁴. In short, in settings where treatments are staggered and dynamic a TWFE regression will estimate treatment effects that combine “good” comparisons between treated and untreated units and “bad” comparisons between just-treated and recently-treated units. Baker et al. (2022) emphasize the importance of these issues in the context of interstate banking deregulation and show that using TWFE instead of a more modern approach can give severely biased results.

³³The nonlinearity of the absolute value function does introduce some bias, which we show in the appendix will actually cause us to *underestimate* the true effect size. Intuitively, because the mean absolute estimation error in $\Delta\hat{c}_{i,t}$ is positive, estimated synchronization is always bounded away from 0 regardless of true synchronization.

³⁴Roth et al. (2023) provides a recent helpful summary of the active and growing literature on this topic.

5.4 Results

Our regressions take this basic form based on Dube et al. (2023):

$$y_{i,j,t+h} - y_{i,j,t-1} = \beta^h \Delta D_{i,j,t} + \delta_t^h + \Gamma \Delta X_{i,j,t-1} + e_{i,j,t}^h$$

where $y_{i,j,t}$ is an outcome such as synchronization for pair $\{i, j\}$, β^h is the h horizon response of outcome y to time t treatment, δ_t^h is a period fixed effect, $\Delta D_{i,j,t}$ is the time t change in treatment status, and $\Delta X_{i,j,t-1}$ are changes in lagged controls. Dube et al. (2023) deal with the problematic comparisons issue by restricting the sample to newly treated observations where $\Delta D_{i,j,t} = 1$ and “clean control” observations that remain untreated h periods out (i.e. where $D_{i,j,t+h} = 0$)³⁵. In our presented results $\Delta X_{i,j,t-1}$ contains the lagged change in economic similarity across the state pair. This measure has been used in Morgan et al. (2004) and Goetz and Gozzi (2022) and is constructed by taking the root mean squared difference in 1-digit industry employment shares between the two states. Our results are robust to its exclusion.

As is usual with a local projection, the coefficient β^h gives the estimated treatment effect at h periods out relative to period $t - 1$. The collection $\{\beta^h\}$ is our estimate of the dynamic effects of treatment. β^h is identified off of the comparison between state-pairs that did deregulate (and therefore begin to integrate) and state pairs that did not. In all of our regressions we use dyad-robust standard errors from Aronow et al. (2017) that allow for unrestricted dependence across any state-pair-years that share a state³⁶.

In figure 8 we then show the responses of consumption and output synchronization to

³⁵Since deregulations are never reversed in our sample (i.e. the treatment is absorbing) the control group for any given state-pair-year contains all of the never-treated pairs and all pairs that are yet to be deregulated h years out.

³⁶Consider two state-pairs i, j and i', j' . A standard two-way clustering approach would allow dependence between pairs where $i = i'$ or $j = j'$. We also allow for dependence when $i = j'$ or $j = i'$. See Cameron and Miller (2014) and Graham (2020) for further discussion.

deregulation. Here, these regressions identify

$$\begin{aligned}\beta_x^h &= E[synch_{i,j,t+h}^x - synch_{i,j,t-1}^x | \Delta X_{i,j,t-1}, D_{i,j,t} = 1] - \\ &E[synch_{i,j,t+h}^x - synch_{i,j,t-1}^x | \Delta X_{i,j,t-1}, D_{i,j,t} = 0]\end{aligned}$$

for $x = cons.$ or $output$. The effect on consumption synchronization is statistically insignificant on impact and rises to about 1.5 percentage points 5 years out. This effect is significant and equal to about half a sample standard deviation or half the initial level of synchronization. Output synchronization rises on impact and stabilizes at about 1 percentage point. The effect is significant on impact but not at all further horizons. As discussed before, we can interpret the difference between these responses as a measure of the *direct* effect of deregulation on risk sharing controlling for the impact of deregulation on output. Given the uncertainty in our estimates we cannot reject that this difference is, in fact, 0. In unreported results, we also test this directly by using $y_{i,j,t} = synch_{i,j,t}^{cons.} - synch_{i,j,t}^{output}$ and find no significant response.

One potential threat to identification is that both consumption and output synchronization appear modestly lower pre-deregulation, which could be a sign of a parallel trends violation. The most likely mechanism is selection into treatment, in which state-pairs that were already becoming more synchronized over time endogenously decide to deregulate. This could cause us to overestimate the true effect size. While the historical evidence presented previously suggests this is not the case, we can address these concerns in our empirical framework as well. As discussed in Dube et al. (2023), if selection into treatment arises from static characteristics, covariates that change over time unpredictably, or lagged changes in our controls then the standard conditional parallel trends assumption will still ensure unbiasedness. If selection into treatment is due to lagged outcome dynamics, then Dube et al. (2023) recommend adding $\Delta y_{i,j,t-k}$ for $k \geq 2$ as controls. We rerun our consumption and output regressions controlling for the first three lags of changes in consumption and output synchronization, respectively, and find similar results with somewhat attenuated point estimates. We include these figures in the appendix.

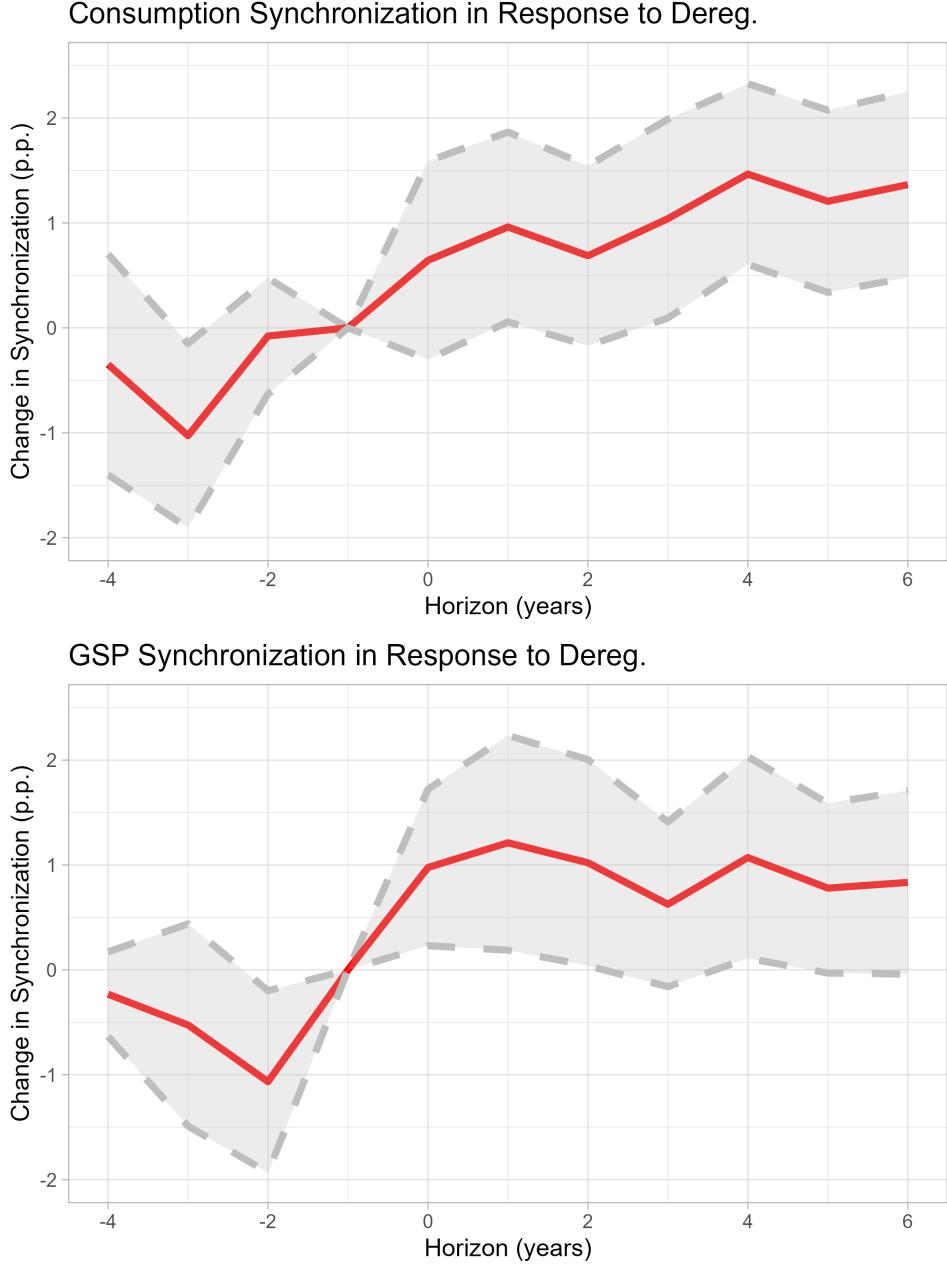


Figure 8: Responses of consumption and output synchronization to banking deregulation. 90% confidence bands shown. Standard errors are dyad-robust (see text for details).

5.5 Discussion

The similarity of the output and consumption responses suggests that there is little effect of integration on risk sharing. To more fully explore this relationship we present additional results in the appendix from an instrumented difference-in-differences specification that al-

lows us to estimate the direct response of synchronization to the level of banking integration using deregulation as an instrument. These results accord with the main ones presented here, suggesting further that output and consumption synchronization respond similarly.

We conclude that this setting provides little evidence that credit markets improve interstate risk sharing. As discussed before, banks facilitate sharing of risks by moving funds across state borders via internal credit markets or national interbank markets. The deregulations considered here enable the expansion of cross-state internal capital markets as measured by $int_{i,j,t}$. Putting aside interbank markets, which are relatively small during our period of interest³⁷, the extent of cross-state banking integration directly determines the capacity for bank-intermediated credit markets to shift risks across states. We therefore interpret the similarity of the consumption and output responses as providing minimal evidence that households are using credit markets to smooth local shocks.

This finding and mechanism contrast with the results of Asdrubali et al. (1996), who find a large but *declining* role for credit markets in risk sharing over our time period. They find that upwards of 35% of output volatility is smoothed by credit markets between 1963 and 1980 but less than 20% is over 1980 to 1990. Their estimates are based on year-by-year cross-sectional regressions of consumption growth on output growth and are subject to the excess smoothness concerns discussed previously. Additionally, as they also mention, their results could be driven by changes in the underlying distribution of shocks to output over time. To the extent that these changes are national or otherwise common between our control and treatment groups our design addresses this issue. More fundamentally, their findings are difficult to reconcile with the low levels of banking integration before deregulation and the increased linkages that develop after 1980.

6 Consumption Fiscal Multipliers

In this section, we estimate dynamic consumption fiscal multipliers using the military spending instruments of Nakamura and Steinsson (2014). In our preferred specification, we find relative consumption multipliers that are positive, close in size to output multipliers, and

³⁷See Allen et al. (2020) for a cross-country comparison and a discussion of the small role of these markets in the U.S. over our time period.

grow substantially over 5 years.

6.1 Empirical Design

There is a long literature in macroeconomics, summarized recently in Ramey (2016) and earlier in Hall (2009), that uses military spending as a source of exogenous variation in total government expenditures and therefore as an instrument to estimate output fiscal multipliers. We follow the strand of this literature exploiting sub-national variation and estimate fiscal multipliers using state-level military spending shocks as measured by two sets of instruments from Nakamura and Steinsson (2014). The first, termed the “interaction” instruments, are constructed by regressing changes in state spending on national spending allowing for differential sensitivities across states. In effect, there is a separate instrument for each state. The fitted values from this estimation are state-specific rescalings of national spending and should be independent of any locally endogenous component of military spending. The second is a simple Bartik instrument, constructed at the state-level by interacting national military spending with fixed, state-specific spending shares. They estimate these shares using the state-level military spending over 1966-1971.

Weak first stages are an important, but surmountable, issue for both the interaction and the Bartik instruments. On our restricted sample of 1970-2006³⁸, both sets of instruments have Kleibergen-Paap heteroskedasticity-robust first stage F-statistics below 5, leaving them vulnerable to the standard array of weak-instrument problems (Andrews et al. (2019)). We do not expect this to be a major issue for two reasons. First, our OLS estimates are close to 0, so our IV estimated multiplier will be, in fact, too small³⁹. Second, we make use of the weak-instrument-robust estimator from Andrews and Armstrong (2017) for the Bartik instrument. Their estimator is unbiased and has good small-sample properties but requires assuming the sign of the first stage coefficient. Given our setting, we are comfortable making that assumption.

Our estimands are dynamic consumption and output multipliers. We adapt the original

³⁸Nakamura and Steinsson (2014) use 1966-2006 and also confront similar weak instrument problems. Our first stages are modestly weaker because we drop several important Vietnam-War-era observations.

³⁹NS assess the importance of weak-instrument issues using simulations and find that they are minor. Since we use a smaller sample than them these findings may not carry over.

NS specification to compute *integral multipliers* as in Ramey and Zubairy (2018) that capture the accumulated change in outcomes over t to $t + h$ in response to the accumulated change in military spending over t to $t + h$, instrumented with a time t Bartik instrument:

$$\begin{aligned} \sum_{j=0}^h \frac{y_{i,t+j} - y_{i,t-1}}{y_{i,t-1}} &= \alpha_i^y + \gamma_t^y + \beta_h^y \sum_{j=0}^h \frac{g_{i,t} - g_{i,t-1}}{y_{i,t-1}} + \Gamma^y X_{i,t-1} + \varepsilon_{i,t} \\ \sum_{j=0}^h \frac{c_{i,t} - c_{i,t-1}}{y_{i,t-1}} &= \alpha_i^c + \gamma_t^c + \beta_h^c \sum_{j=0}^h \frac{g_{i,t} - g_{i,t-1}}{y_{i,t-1}} + \Gamma^c X_{i,t-1} + \eta_{i,t}^c \\ \sum_{j=0}^h \frac{g_{i,t} - g_{i,t-1}}{y_{i,t-1}} &= \alpha_i^z + \gamma_t^z + \Pi z_{i,t} + \Gamma^z X_{i,t-1} + \nu_{i,t} \end{aligned}$$

where $y_{i,t}$ is real output per capita in state i , $c_{i,t}$ is real estimated consumption per capita, $g_{i,t}$ is real per capita military spending, X is a collection of time $t - 1$ or earlier controls, and $z_{i,t}$ is either the set of interaction instruments or the single Bartik instrument. We use the national consumer price index (CPI) to convert from nominal to real quantities. Previous research has used the state-level price series constructed in Del Negro (2002), which is constructed from official city-level price indices and cost-of-living estimates from a private firm. Similar to private-sector consumption estimates, the measurement error in these series is largely unknown, so we make use of national numbers⁴⁰. Our regressors and regressands are normalized by lagged output following Hall (2009). Our confidence sets are constructed by inverting an Anderson-Rubin test and by using the weak-IV-robust confidence sets from Andrews and Armstrong (2017) for the interactions and Bartik instruments, respectively.

Several features of our specification are also chosen to address excess smoothness concerns as discussed in the evaluation section. First, we drop from our sample states the five states without sales taxes due to the limited information we have for them in the earlier parts of our sample. Additionally, we control for two lags of our outcome variables. We discuss these choices further in the appendix.

⁴⁰Hazell et al. (2022) constructs state-specific price indices using BLS microdata. Unfortunately, these series' coverage across states is limited and begins in 1978.

6.2 Results and Discussion

Our estimated dynamic fiscal multipliers are shown in figure 9. We use the Bartik instrument and the Andrews and Armstrong (2017) estimator to estimate the sequences $\{\beta_h^y\}_{h=0}^5$ and $\{\beta_h^c\}_{h=0}^5$ and their 90% confidence intervals. The results using the interactions instruments, unreported here, are similar but less precise. In all regressions the standard errors are Driscoll-Kraay with two lags to account for cross-sectional dependence and serial correlation in the errors. Both our output and consumption multipliers are imprecisely estimated and positive on impact, growing over time, and quite large at five years out. Our output estimates at $h = 2$ are similar in magnitude to those reported in Nakamura and Steinsson (2014)⁴¹. Our estimates and confidence bands allow us to reject consumption multipliers below 1 at five years out.

We first compare our estimates to those from two related papers. First, Dupor et al. (2023) similarly estimate regional fiscal multipliers for a wide basket of consumption goods. Their study focuses on the Great Recession and exploits regional variation in American Recovery and Reinvestment Act spending. They measure consumption using the Nielsen retail scanner dataset, which covers many common nondurable retail purchases, and car loan data from a credit bureau. Dupor et al. (2023) then use the average relationship between Nielsen-type nondurables and other consumption categories in the CEX to infer multipliers for larger categories of durables, nondurables and services. Roughly matching their categories to our notion of retail spending gives a four-year consumption fiscal multiplier around 0.30-0.35 compared to our four-year point estimate of about 2. We can clearly reject values that low given our estimated confidence intervals and interpret this as moderate evidence that the consumption fiscal multiplier over our sample is larger than that estimated by Dupor et al. (2023). Given that output multipliers are typically estimated to be larger in times of slack⁴², it is notable that our responses, which are estimated across many expansions and contractions, are larger than those based on evidence from the Great Recession. One potential explanation is that the subset of the basket that Dupor et al. (2023)

⁴¹We present static multipliers estimated using the NS specification in the appendix.

⁴²Ramey and Zubairy (2018) provides an overview.

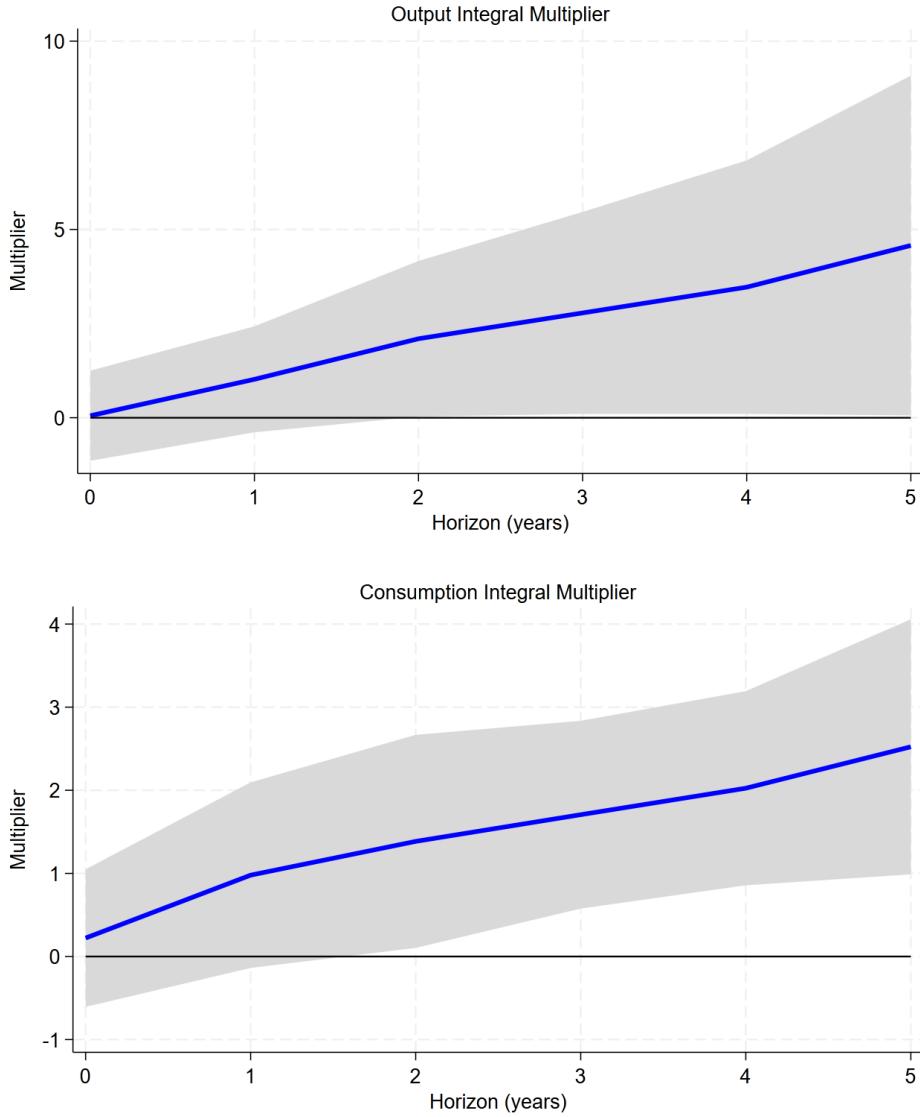


Figure 9: Estimates of output and consumption integral fiscal multipliers using Nakamura and Steinsson (2014) military spending shocks and the Andrews et al. (2019) weak-IV-robust estimator. 90% confidence bands shown. Standard errors are Driscoll-Kraay.

can directly measure responds differently than overall retail spending in ways not captured by the average relationships seen in the CEX. Additionally, it is possible that difference in expected persistence between ARRA and military contract spending could generate this discrepancy.

A second point of comparison is Chen (2019), who estimates a long-horizon variant of the original NS specification using retail employment as a proxy for real consumption. His is one

of few papers to estimate dynamic output fiscal multipliers using regional data and is, to our knowledge, the only one to do so for consumption. He estimates a 5-year response of retail employment to a military spending shock of about 3 and an output response of about 2.5 using a somewhat similar long difference multiplier specification and the interactions IV. We view his findings as corroborating evidence that consumption and output regional multipliers are large in the long-run. Additionally, the similarity between his real consumption responses and ours suggests that our estimates are not substantially biased upwards due to local price responses to fiscal shocks.

Lastly, we briefly connect our results to existing models of regional and aggregate fiscal multipliers. In many models of regional fiscal multipliers, such as the standard New Keynesian one in Farhi and Werning (2016) or the HANK one in Dupor et al. (2023), short run regional output and consumption multipliers are typically small and translate to smaller or even negative aggregate multipliers depending on the fiscal authority’s financing strategy or the response of monetary policy to fiscal stimulus. Additionally, these models cannot generate the large, sustained medium run responses that we observe due to a lack of internal propagation mechanisms. Chen (2019) proposes one possible solution: immigration in response to fiscal stimulus. Looking at the response of state populations to the NS fiscal shocks he finds evidence for this channel. Explicitly modeling the response of population or other “stocks” to fiscal shocks—via e.g. capital inflows or increased firm creation—could potentially close the gap between existing models and our empirical findings.

7 Conclusion

In this paper we have constructed a new set of estimates for state-level retail consumption. We built our estimates off of the information contained in various direct measures of retail spending, including newly digitized state sales tax records. Our statistical approach accommodated the various features of our input datasets, including missing data, varying levels of measurement error, and aggregated observations. Our estimates performed sensibly when compared to national PCE and leading alternative measures of consumption. We applied our estimates first to the study of interstate risk sharing via banking linkages. We designed our empirical approach to address concerns about the excess smoothness of filtered consumption

estimates and found that increasing banking integration did not increase risk sharing. In our interpretation, this implies that household do not use credit markets to do ex-post smoothing of local shocks. We then turned our attention to consumption fiscal multipliers. Using state-specific military spending shocks, we showed that consumption fiscal multipliers are typically large and close to output multipliers. Our estimates are larger than those based on private-sector consumption data from the Great Recession and provide additional moments for models of regional fiscal multipliers to match.

We view this paper as the starting point for two types of future work. First, further refinements of our consumption estimates are possible. While our model is annual, much of the data we use is available at higher frequency and could be used to estimate a quarterly consumption series. Additionally, detailed sectoral breakdowns are available in the economic census and in some state sales tax records and MRTS reports. These could be used to construct durable and nondurable consumption series for a subset of states. Our estimates can also be extended back to at least 1967, as the economic census and some states' MRTS and sales tax series continue into the 1960s. Second, there are many additional natural applications of state-level consumption data. We look forward to future work that utilizes our estimates to answer a variety of interesting economic questions.

References

- Chadi S Abdallah and William D Lastrapes. Home Equity Lending and Retail Spending: Evidence from a Natural Experiment in Texas. *American Economic Journal: Macroeconomics*, 4(4):94–125, October 2012. ISSN 1945-7707, 1945-7715. doi: 10.1257/mac.4.4.94. URL <https://pubs.aeaweb.org/doi/10.1257/mac.4.4.94>.
- Mario Alloza, Jesús Gonzalo, and Carlos Sanz. Dynamic Effects of Persistent Shocks, December 2019. URL <https://papers.ssrn.com/abstract=3508672>.
- Dean F Amel. State laws affecting the geographic expansion of commercial banks. 2000.
- Isaiah Andrews and Timothy B. Armstrong. Unbiased instrumental variables estimation under known first-stage sign. *Quantitative Economics*, 8(2):479–503, 2017. ISSN 1759-7331. doi: 10.3982/QE700. URL <https://onlinelibrary.wiley.com/doi/abs/10.3982/QE700>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE700>.
- Isaiah Andrews, James H. Stock, and Liyang Sun. Weak Instruments in Instrumental Variables Regression: Theory and Practice. *Annual Review of Economics*, 11(1):727–753, August 2019. ISSN 1941-1383, 1941-1391. doi: 10.1146/annurev-economics-080218-025643. URL <https://www.annualreviews.org/doi/10.1146/annurev-economics-080218-025643>.
- Juan Antolin-Diaz, Thomas Drechsel, and Ivan Petrella. Tracking the Slowdown in Long-Run GDP Growth. *The Review of Economics and Statistics*, 99(2):343–356, May 2017. ISSN 0034-6535, 1530-9142. doi: 10.1162/REST_a_00646. URL <https://direct.mit.edu/rest/article/99/2/343-356/58399>.
- Peter M. Aronow, Cyrus Samii, and Valentina A. Assenova. Cluster-Robust Variance Estimation for Dyadic Data. *Political Analysis*, 23(4):564–577, January 2017. ISSN 1047-1987, 1476-4989. doi: 10.1093/pan/mpv018. URL <https://www.cambridge.org/core/journals/political-analysis/article/abs/clusterrobust-variance-estimation-for-dyadic-data/D43E12BF35240100C7A4ED3C28912C95>. Publisher: Cambridge University Press.

S. Borağan Aruoba, Francis X. Diebold, Jeremy Nalewaik, Frank Schorfheide, and Dongho Song. Improving G D P measurement: A measurement-error perspective. *Journal of Econometrics*, 191(2):384–397, April 2016. ISSN 03044076. doi: 10.1016/j.jeconom.2015.12.009. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304407615002857>.

P. Asdrubali, B. E. Sorensen, and O. Yosha. Channels of Interstate Risk Sharing: United States 1963-1990. *The Quarterly Journal of Economics*, 111(4):1081–1110, November 1996. ISSN 0033-5533, 1531-4650. doi: 10.2307/2946708. URL <https://academic.oup.com/qje/article-lookup/doi/10.2307/2946708>.

Christian Awuku-Budu, Ledia Guci, Christopher A Lucas, and Carol A Robbins. A First Look at Experimental Personal Consumption Expenditures by State. 2013.

Andrew C. Baker, David F. Larcker, and Charles C.Y. Wang. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395, May 2022. ISSN 0304405X. doi: 10.1016/j.jfineco.2022.01.004. URL <https://linkinghub.elsevier.com/retrieve/pii/S0304405X22000204>.

Michael Cai, Marco Del Negro, Edward Herbst, Ethan Matlin, Reca Sarfati, and Frank Schorfheide. Online estimation of DSGE models. *The Econometrics Journal*, 24(1):C33–C58, March 2021. ISSN 1368-4221, 1368-423X. doi: 10.1093/ectj/utaa029. URL <https://academic.oup.com/ectj/article/24/1/C33/5909595>.

A Colin Cameron and Douglas L Miller. Robust Inference for Dyadic Data. 2014.

Vince W Chen. Fiscal multipliers and regional reallocation, 2019.

Gabriel Chodorow-Reich. Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned? *American Economic Journal: Economic Policy*, 11(2):1–34, May 2019. ISSN 1945-7731, 1945-774X. doi: 10.1257/pol.20160465. URL <https://pubs.aeaweb.org/doi/10.1257/pol.20160465>.

Gabriel Chodorow-Reich. Regional data in macroeconomics: Some advice for practitioners. *Journal of Economic Dynamics and Control*, 115:103875, June 2020. ISSN 0165-1889. doi:

10.1016/j.jedc.2020.103875. URL <https://www.sciencedirect.com/science/article/pii/S0165188920300440>.

Peter K. Clark. The Cyclical Component of U. S. Economic Activity. *The Quarterly Journal of Economics*, 102(4):797, November 1987. ISSN 00335533. doi: 10.2307/1884282. URL <https://academic.oup.com/qje/article-lookup/doi/10.2307/1884282>.

Marco Del Negro. Asymmetric shocks among US states. *Journal of International Economics*, 56(2):273–297, 2002. ISBN: 0022-1996 Publisher: Elsevier.

Yuliya Demyanyk, Charlotte Oستergaard, and Bent E. Sørensen. U.S. Banking Deregulation, Small Businesses, and Interstate Insurance of Personal Income. *The Journal of Finance*, 62(6):2763–2801, 2007. ISSN 0022-1082. URL <https://www.jstor.org/stable/4622353>. Publisher: [American Finance Association, Wiley].

Arindrajit Dube, Daniele Girardi, Óscar Jordà, and Alan M. Taylor. A Local Projections Approach to Difference-in-Differences Event Studies, April 2023. URL <https://www.nber.org/papers/w31184>.

Bill Dupor, Marios Karabarounis, Marianna Kudlyak, and M Saif Mehkari. Regional Consumption Responses and the Aggregate Fiscal Multiplier. *Review of Economic Studies*, page rdad007, April 2023. ISSN 0034-6527, 1467-937X. doi: 10.1093/restud/rdad007. URL <https://academic.oup.com/restud/advance-article/doi/10.1093/restud/rdad007/7103489>.

E. Farhi and I. Werning. Chapter 31 - Fiscal Multipliers
We thank the editors John Taylor and Harald Uhlig for detailed comments, as well as suggestions and comments by Gabriel Chodorow-Reich, Jon Steinsson, and Michael Weber.: Liquidity Traps and Currency Unions. In John B. Taylor and Harald Uhlig, editors, *Handbook of Macroeconomics*, volume 2, pages 2417–2492. Elsevier, January 2016. doi: 10.1016/bs.hesmac.2016.06.006. URL <https://www.sciencedirect.com/science/article/pii/S1574004816300192>.

Leland Farmer, Emi Nakamura, and Jón Steinsson. Learning About the Long Run. Technical

Report w29495, National Bureau of Economic Research, Cambridge, MA, November 2021.
URL <http://www.nber.org/papers/w29495.pdf>.

Thomas A Garrett, Rubn Hernndez-Murillo, and Michael T Owyang. Does Consumer Sentiment Predict Regional Consumption? *Federal Reserve Bank of St. Louis Review*, 2005.

Martin R. Goetz and Juan Carlos Gozzi. Financial integration and the co-movement of economic activity: Evidence from U.S. states. *Journal of International Economics*, 135: 103561, March 2022. ISSN 0022-1996. doi: 10.1016/j.jinteco.2021.103561. URL <https://www.sciencedirect.com/science/article/pii/S0022199621001410>.

Bryan S. Graham. Dyadic regression. In *The Econometric Analysis of Network Data*, pages 23–40. Elsevier, 2020. ISBN 978-0-12-811771-2. doi: 10.1016/B978-0-12-811771-2.00008-0. URL <https://linkinghub.elsevier.com/retrieve/pii/B9780128117712000080>.

Adam Guren, Alisdair McKay, Emi Nakamura, and Jón Steinsson. What Do We Learn from Cross-Regional Empirical Estimates in Macroeconomics? *NBER Macroeconomics Annual*, 35:175–223, May 2021. ISSN 0889-3365, 1537-2642. doi: 10.1086/712321. URL <https://www.journals.uchicago.edu/doi/10.1086/712321>.

Robert E. Hall. By How Much Does GDP Rise if the Government Buys More Output?, November 2009. URL <https://www.nber.org/papers/w15496>.

James D. Hamilton. *Time series analysis*. Princeton University Press, Princeton, N.J, 1994. ISBN 978-0-691-04289-3.

Edward Herbst and Frank Schorfheide. Sequential Monte Carlo Sampling for Dsge Models. *Journal of Applied Econometrics*, 29(7):1073–1098, 2014. ISSN 1099-1255. doi: 10.1002/jae.2397. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2397>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.2397>.

Edward Herbst and Frank Schorfheide. *Bayesian Estimation of DSGE Models*. Princeton University Press, December 2016. ISBN 978-1-4008-7373-9. doi: 10.1515/9781400873739. URL <https://www.degruyter.com/document/doi/10.1515/9781400873739/html>.

Gregory D Hess and Kwanho Shin. Intranational business cycles in the United States. *Journal of International Economics*, 44(2):289–313, April 1998. ISSN 00221996. doi: 10.1016/S0022-1996(97)00032-9. URL <https://linkinghub.elsevier.com/retrieve/pii/S0022199697000329>.

Sally Hudson, Peter Hull, and Jack Lieberson. Interpreting Instrumented Difference-in-Differences, 2017. URL <http://www.mit.edu/~liebers/DDIV.pdf>.

J. Jayaratne and P. E. Strahan. The Finance-Growth Nexus: Evidence from Bank Branch Deregulation. *The Quarterly Journal of Economics*, 111(3):639–670, August 1996. ISSN 0033-5533, 1531-4650. doi: 10.2307/2946668. URL <https://academic.oup.com/qje/article-lookup/doi/10.2307/2946668>.

Sebnem Kalemli-Ozcan, Elias Papaioannou, and José-Luis Peydró. Financial Regulation, Financial Globalization, and the Synchronization of Economic Activity. *The Journal of Finance*, 68(3):1179–1228, 2013. ISSN 1540-6261. doi: 10.1111/jofi.12025. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12025>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12025>.

Tim A. Kroencke. Asset Pricing without Garbage. *The Journal of Finance*, 72(1):47–98, 2017. ISSN 1540-6261. doi: 10.1111/jofi.12438. URL <http://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12438>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12438>.

R. S. Kroszner and P. E. Strahan. What Drives Deregulation? Economics and Politics of the Relaxation of Bank Branching Restrictions. *The Quarterly Journal of Economics*, 114(4):1437–1467, November 1999. ISSN 0033-5533, 1531-4650. doi: 10.1162/003355399556223. URL <https://academic.oup.com/qje/article-lookup/doi/10.1162/003355399556223>.

R. S. Kroszner and Philip E Strahan. *Regulation and deregulation of the US banking industry: causes, consequences, and implications for the future*. Economic regulation and its reform: what have we learned? University of Chicago Press, 2014. ISBN 978-

0-226-13802-2 978-0-226-13816-9. doi: 10.7208/chicago/9780226138169.001.0001. URL <https://www.bibliovault.org/BV.landing.epl?ISBN=9780226138169>.

Atif Mian, Amir Sufi, and Emil Verner. How Does Credit Supply Expansion Affect the Real Economy? The Productive Capacity and Household Demand Channels. *The Journal of Finance*, 75(2):949–994, 2020. ISSN 1540-6261. doi: 10.1111/jofi.12869. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12869>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jofi.12869>.

John L Mikesell and Sharon N Kioko. The retail sales tax in a new economy. In *7th Annual Municipal Finance Conference, Hutchins Center on Fiscal and Monetary Policy, Washington, DC [www. brookings. edu/wp-content/uploads/2018/04/Mikesell-Kioko1. pdf]*, 2018.

Donald P Morgan, Bertrand Rime, and Philip E Strahan. Bank Integration and State Business Cycles. *Quarterly Journal of Economics*, page 30, 2004.

James C. Morley, Charles R. Nelson, and Eric Zivot. Why Are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP So Different? *Review of Economics and Statistics*, 85(2):235–243, May 2003. ISSN 0034-6535, 1530-9142. doi: 10.1162/003465303765299765. URL <https://direct.mit.edu/rest/article/85/2/235-243/57373>.

Emi Nakamura and Jón Steinsson. Fiscal Stimulus in a Monetary Union: Evidence from US Regions. *American Economic Review*, 104(3):753–792, March 2014. ISSN 0002-8282. doi: 10.1257/aer.104.3.753. URL <https://pubs.aeaweb.org/doi/10.1257/aer.104.3.753>.

Maurice Obstfeld. Are Industrial-Country Consumption Risks Globally Diversified? Technical report, 1994. URL https://www.nber.org/system/files/working_papers/w4308/w4308.pdf.

Maxim Pinkovskiy and Xavier Sala-i Martin. Lights, Camera . . . Income! Illuminating the National Accounts-Household Surveys Debate *. *The Quarterly Journal of Economics*, 131(2):579–631, May 2016. ISSN 0033-5533. doi: 10.1093/qje/qjw003. URL <https://doi.org/10.1093/qje/qjw003>.

V.A. Ramey. Macroeconomic Shocks and Their Propagation. In *Handbook of Macroeconomics*, volume 2, pages 71–162. Elsevier, 2016. ISBN 978-0-444-59487-7. doi: 10.1016/bs.hesmac.2016.03.003. URL <https://linkinghub.elsevier.com/retrieve/pii/S1574004816000045>.

Valerie A. Ramey and Sarah Zubairy. Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data. *Journal of Political Economy*, 126(2):850–901, April 2018. ISSN 0022-3808. doi: 10.1086/696277. URL <https://www.journals.uchicago.edu/doi/full/10.1086/696277>. Publisher: The University of Chicago Press.

Patrick Rebeschini and Ramon Van Handel. Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25(5), October 2015. ISSN 1050-5164. doi: 10.1214/14-AAP1061. URL <https://projecteuclid.org/journals/annals-of-applied-probability/volume-25/issue-5/Can-local-particle-filters-beat-the-curse-of-dimensionality/10.1214/14-AAP1061.full>.

Jonathan Roth, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, August 2023. ISSN 0304-4076. doi: 10.1016/j.jeconom.2023.03.008. URL <https://www.sciencedirect.com/science/article/pii/S0304407623001318>.

Alexi Savov. Asset Pricing with Garbage. *The Journal of Finance*, 66(1):177–201, February 2011. ISSN 00221082. doi: 10.1111/j.1540-6261.2010.01629.x. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2010.01629.x>.

Frank Schorfheide, Dongho Song, and Amir Yaron. Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach. *Econometrica*, 86(2):617–654, 2018. ISSN 0012-9682. doi: 10.3982/ECTA14308. URL <https://www.econometricsociety.org/doi/10.3982/ECTA14308>.

David W Wilcox. The Construction of U.S. Consumption Data: Some Facts and Their Implications for Empirical Work. *American Economic Review*, 82(4):922–941, 1992.

Xia Zhou. *Essays on US state-level financial wealth data and consumption data*. The Johns Hopkins University, 2010.

Appendix

A Additional Estimation Details

See figure 10 for plots of the posterior distributions of the main parameters.

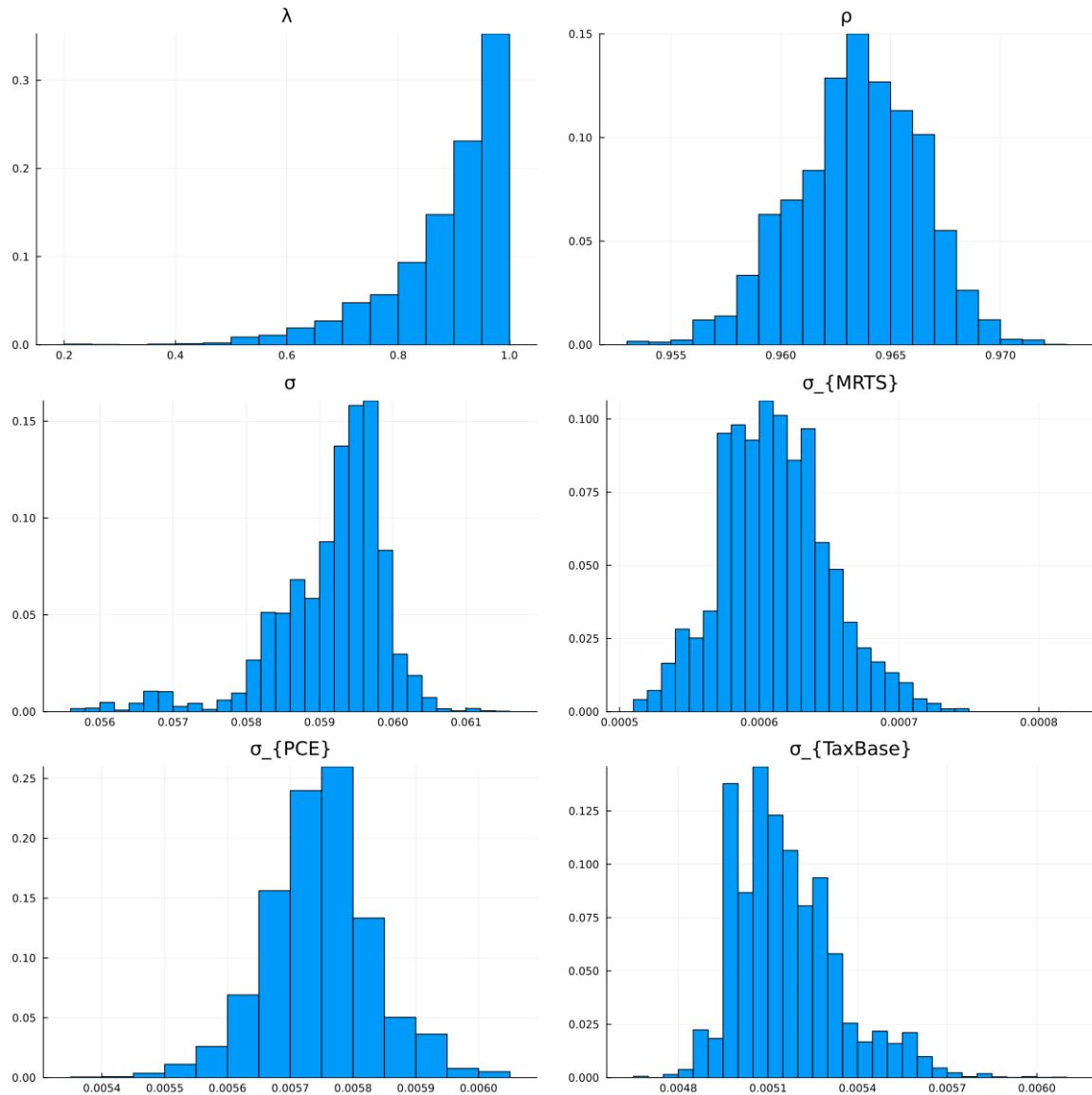


Figure 10: Unweighted posterior distributions for main parameters

B Additional Comparisons between Estimates and Alternatives

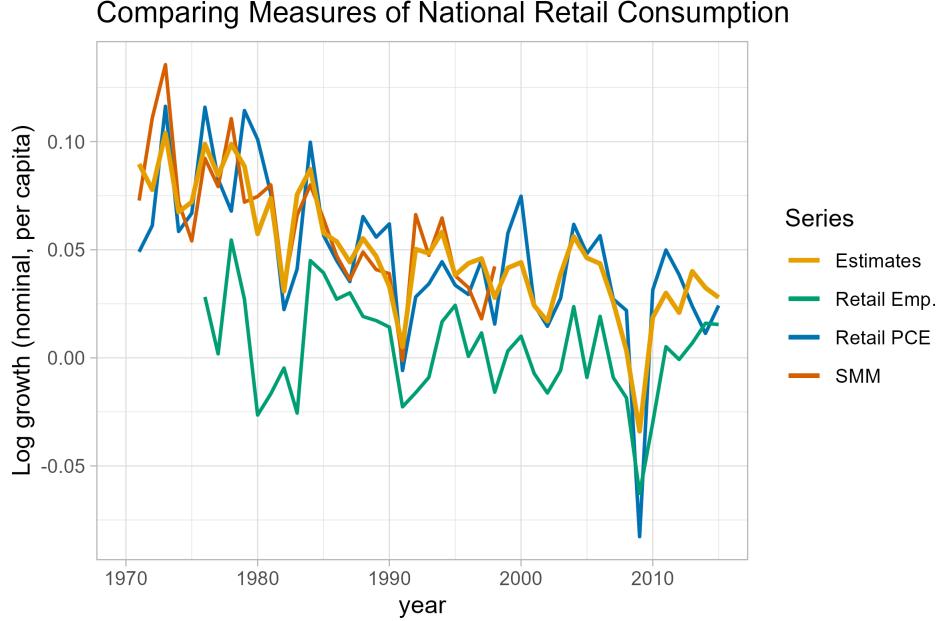


Figure 11: Nationally aggregated estimated versus alternatives

C More on Excess Smoothness

In this section we lay out a simple two-observable model of excess smoothness in filtered quantities and discuss how excess smoothness can impact other series as well.

C.1 A Simple Model of Excess Smoothness

To understand why there is excess smoothness in our estimates, it is helpful to consider a simple single-region version of our model and the properties of its Kalman filtered estimates⁴³. Let true consumption c_t evolve as a random walk and y_t^A and y_t^B be two noisy signals for c_t that have independent measurement errors:

$$\begin{aligned} c_t &= c_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \\ y_t^A &= c_t + \eta_t^A, \quad \eta_t^A \sim N(0, \sigma_A^2) \\ y_t^B &= c_t + \eta_t^B, \quad \eta_t^B \sim N(0, \sigma_B^2) \end{aligned}$$

After a few initial periods⁴⁴, the estimates $\hat{c}_t = E[c_t | y_{1:t}^A, y_{1:t}^B]$ from the Kalman filter evolve according to this law of motion:

$$\hat{c}_t = (1 - K^A - K^B)\hat{c}_{t-1} + K^A y_t^A + K^B y_t^B \quad (1)$$

⁴³The equivalent derivations for the smoothed (two-sided) estimates carry the similar intuition but are less clear. Kroencke (2017) provides a similar exposition for the case with a single observable and stochastic volatility.

⁴⁴Kalman filters have a initial “burn-in” period after which the Kalman gain converges towards its steady-state value. We work with a steady-state filter for clarity.

Time t estimates of consumption are a weighted average of the observations y_t^A and y_t^B and the model's forecast for c_t based on the last period's information, which is \hat{c}_{t-1} since c_t is a random walk. The weights on these three components are based on K^A and K^B , the elements of the steady-state Kalman gain corresponding to the variables y^A and y^B , respectively. These Kalman gains measure the "informativeness" of each signal and their derivation and precise forms are provided below.

The degree of excess smoothness in the estimates is proportional to the measurement error in the observed data. Intuitively, each series' Kalman gain is larger when that series is less noisy in either an absolute or comparative sense. In the limiting case where either $\sigma_A = 0$ or $\sigma_B = 0$ we have that the corresponding Kalman gain converges to 1 and $\hat{c}_{i,t} = c_{i,t}$ exactly with no excess smoothness. Conversely, taking $\sigma_A, \sigma_B \rightarrow \infty$ implies $K^A, K^B \rightarrow 0$ and that the $\hat{c}_{i,t}$ are perfectly smooth.

Furthermore, the estimates are smoother when data is missing. Suppose, without loss of generality, that a single period of y_t^A is missing. The Kalman filter accommodates this by setting $K^A = 0$ and computing a new $K^B = \tilde{K}^B$ for this period. \tilde{K}^B is the Kalman gain that arises from a model where only y_t^B is observed⁴⁵. We show below that $1 - \tilde{K}^B > 1 - K^A - K^B \forall \sigma_A, \sigma_B > 0$, so that the consumption estimates are necessarily smoother when y_t^A is missing. This can be seen visually in the estimates for Rhode Island in figure 4. Between 1988 and 1995, when no state-specific information is available, consumption estimates are much smoother than they are otherwise. In general, states-years with fewer observed data points will be smoother than state-years with more data.

This excess smoothness in estimated consumption can introduce bias in applications of the data. To fix ideas, consider a standard risk sharing regression along the lines of Obstfeld (1994) or Asdrubali et al. (1996) where filtered consumption growth is regressed on output growth o_t : $\Delta\hat{c}_t = \alpha + \beta\Delta o_t + \epsilon_{i,t}$. The researcher running this regression has the estimand $\beta = \text{Cov}(\Delta c_t, \Delta o_t)/V(\Delta o_t)$ in mind, but if she uses *estimated* consumption she will instead recover:

$$\beta = \text{Cov}(\Delta\hat{c}_t, \Delta o_t)/V(\Delta o_t) \quad (2)$$

$$= \underbrace{(K^A + K^B)}_{\text{attenuated true coefficient}} \frac{\text{Cov}(\Delta c_t, \Delta o_t)}{\text{Var}(\Delta o_t)} + \underbrace{\sum_{j=1}^{\infty} (1 - K^A - K^B)^j (K^A + K^B)}_{\text{bias from higher order covariances}} \frac{\text{Cov}(\Delta c_{t-j}, \Delta o_t)}{\text{Var}(\Delta o_t)} \quad (3)$$

where we have used that time t estimates can be written as a function of all past observations $\hat{c}_{i,t} = \sum_{j=0}^{\infty} (1 - K^A - K^B)^j (K^A y_{t-j}^A + K^B y_{t-j}^B)$, that $y_t^i = c_t + \eta_t^i$ with η_t^i iid measurement error for $i = A, B$, and the first difference of equation 1. The bias introduced by using consumption estimates comes from the presence of measurement error in each period's observations and the model's use of current and past data to infer c_t . The magnitude of this bias is decreasing in the total informativeness of the data $K^A + K^B$. In practice, since the informativeness of the data can vary across states and years, the extent of the issue will vary as well. Applications where estimated consumption is used as a regressor will also be biased for the same reasons.

⁴⁵Alternatively, it can be calculated by taking the measurement error of series y^A to infinity: $\tilde{K}^B = \lim_{\sigma_A \rightarrow \infty} K^B$. In either case we use the steady-state values of the Kalman gains.

Derivations

Using the model laid out in 3, define the following matrices:

$$\begin{aligned} T &= [1], Z = [1 \ 1]', R = [1], Q = [\sigma^2], \\ E &= \begin{bmatrix} \sigma_A^2 & \\ & \sigma_B^2 \end{bmatrix} \\ P &= TP(T - KZ)' + RQR', V = ZPZ' + E \\ K &= TP'Z'V^{-1} \end{aligned}$$

Under the standard Kalman filter recursions K is the Kalman gain, the estimated variance of true consumption c_t converges to a single element matrix $P = [p^2]$ such that $P = TP(T - KZ)' + RQR'$, and the estimated variance of the observables $y_t = [y_t^A y_t^B]'$ converges to $V = ZPZ' + E$. In our non-stationary setting, this limiting value P is a deterministic function of our initial guess P_0 and is not pinned down by other model parameters. This does not impact the results presented here. Simplifying the expression for K gives

$$K = [K^A \ K^B] = \begin{bmatrix} \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2 + \sigma_A^2 \sigma_B^2 / p^2} & \frac{\sigma_A^2}{\sigma_A^2 + \sigma_B^2 + \sigma_A^2 \sigma_B^2 / p^2} \end{bmatrix}$$

It is clear then that $K^A, K^B \in (0, 1)$ and $(K^A + K^B) < 1$. If we take the limits $\sigma_A \rightarrow 0$ and $\sigma_A \rightarrow \infty$ we get $K = [1 \ 0]$ and $K = [0 \ \frac{p^2}{p^2 + \sigma_B^2}]$, respectively. The first limit corresponds to the case in which y^A provides a perfect measurement of c_t and the second the case in which it provides no information at all. Note that this second case is equivalent to a model with only a single observable. The same holds for the corresponding σ_B limits. Additionally, note that $\partial K^i / \partial \sigma^i < 0$ and $\partial K^i / \partial \sigma^j > 0$ for $i, j \in \{A, B\}$, so that the weight placed by the filter on a given observable is decreasing in that variable's measurement error variance and increasing in the other variable's measurement error variance.

Lastly, we can show that more weight is placed on the data when more sources of data are available. The standard Kalman filter recursions allow us to write each period's estimated consumption as the weighted average of the previous period's estimate and newly available data with weights determined by the Kalman gains: $\hat{c}_t = (1 - K^A - K^B)\hat{c}_{t-1} + K^A y_t^A + K^B y_t^B$. A period in which y_t^A is missing is equivalent to a situation in which its measurement error is infinite. Let $\tilde{K}^B = \lim_{\sigma_A \rightarrow \infty} K_B$. We aim to show that $1 - \tilde{K}^B > 1 - K^A - K^B \forall \sigma^A, \sigma^B > 0$, so that the consumption estimates are necessarily smoother when y_t^A is missing. Towards a contradiction, assume $1 - \tilde{K}^B < 1 - K^A - K^B$. Then,

$$\begin{aligned} 1 - \frac{\sigma_A^2 + \sigma_B^2}{\sigma_A^2 + \sigma_B^2 + \sigma_A^2 \sigma_B^2 / p^2} &> 1 - \frac{p^2}{p^2 + \sigma_A^2} = 1 - \frac{1}{1 + \sigma_A^2 / p^2} \implies \\ \sigma_A^2 + \sigma_B^2 + \sigma_A^2 \sigma_B^2 / p^2 &> \sigma_A^4 / p^2 + \sigma_A^2 + \sigma_B^2 + \sigma_A^2 \sigma_B^2 / p^2 \implies \\ 0 &> \sigma_A^4 / p^2 \end{aligned}$$

which must be false since $p^2 > 0$ and $\sigma_A^2 \geq 0$.

C.2 Excess Smoothness Elsewhere

These excess smoothness concerns are an issue for a variety of data series beyond just our estimates. Commonly used series such as PCE and SMM are also products of filtering

processes⁴⁶ and are subject to the similar excess smoothness concerns as our estimates. On the other hand, “raw” unfiltered series such as the MRTS or the sales tax base are noisy but not excessively smoothed and therefore can be used without introducing bias. Kroencke (2017) illustrates this by comparing official PCE with municipal garbage volumes, a noisy measure of true consumption used by Savov (2011). He finds that using PCE instead of garbage as the regressand in asset-pricing regressions leads to substantial attenuation issues due to excess smoothness. In unreported results, we find similar attenuation issues when we estimate the risk sharing regression from above using our estimates or the SMM series versus the MRTS data on the subset of states and years where the MRTS is available⁴⁷. With these concerns in mind, there are two simple ways that users of our estimates can minimize the impact of excess smoothness.

D Additional Results for Banking Deregulation

D.1 IV Results

Our endogenous variable is banking integration and our instrument is an indicator variable marking whether a given state pair has been deregulated. For ease of interpretation, we make use of a standardized measure of banking integration in the following results.

While the reduced form results give the impact of deregulation on synchronization, we can also estimate the causal relationship between banking integration and synchronization directly. To do so, we estimate:

$$\begin{aligned} y_{i,j,t+h} - y_{i,j,t-1} &= \beta_{DDIV}^h (int_{i,j,t+h} - int_{i,j,t-1}) + \delta_t^h + \Gamma \Delta X_{i,j,t-1} + e_{i,j,t}^h \\ int_{i,j,t+h} - int_{i,j,t-1} &= \gamma^h \Delta D_{i,j,t} + \alpha_t^h + \Lambda \Delta X_{i,j,t-1} + v_{i,j,t}^h \end{aligned}$$

where the estimand for the coefficient β_{DDIV}^h is:

$$\begin{aligned} \beta_{DDIV}^h &= E[y_{i,j,t+h} - y_{i,j,t-1} | \Delta X, int_{i,j,t+h} - int_{i,j,t-1} = 1] - \\ &\quad E[y_{i,j,t+h} - y_{i,j,t-1} | \Delta X, int_{i,j,t+h} - int_{i,j,t-1} = 0] \end{aligned}$$

Each stage is estimated using the difference-in-difference setup of Dube et al. (2023). This instrumented difference-in-difference specification correctly estimates β_{DDIV}^h under two sets of assumptions. First, we need the usual exogeneity and relevance assumptions for the instrument. Second, we need the no anticipation and parallel trends assumptions to hold in *both* the first stage and reduced forms. Hudson et al. (2017) discusses these assumptions and the estimand of an instrumented difference-in-differences design further.

β_{DDIV}^h estimates the change of y over h periods in response to a 1 standard deviation increase in integration over h periods. We interpret this as the impact of a policy intervention

⁴⁶Though PCE and SMM, to our knowledge, are not estimated using a Kalman filter, they are both constructed with similar goals in mind. Kroencke (2017) presents evidence that national PCE is well modeled as the output from a simple one-variable Kalman filtering model.

⁴⁷Here, attenuation causes the researcher to *overestimate* the extent of risk sharing. The difference between raw and filtered consumption measures may therefore explain some of the differences between Hess and Shin (1998) (who use the MRTS and find little interstate risk sharing) and Asdrubali et al. (1996) (who use SMM and find much interstate risk sharing).

that increases integration by one standard deviation from $t - 1$ to $t + h$, having directly instrumented for this long-run change. Another way to view them is as a generalization of the within-period instrumented difference-in-difference estimates to a dynamic setting.

In figure 12 we show the first stage response of standardized banking integration to deregulation. Here, the effect identified is:

$$\beta_{\text{FirstStage}}^h = E[int_{i,j,t+h} - int_{i,j,t-1} | \Delta X_{i,j,t-1}, D_{i,j,t} = 1] - E[int_{i,j,t+h} - int_{i,j,t-1} | \Delta X_{i,j,t-1}, D_{i,j,t} = 0]$$

The identification of this treatment effect depends on treatment and control groups evolving along parallel trends, conditional on our controls, and no response of banking integration in anticipation of treatment. Our pre-treatment effects are tightly estimated zeros, providing no evidence against the parallel trends assumption. Post-treatment, the level of banking integration grows steadily over time. Our estimated dynamic effects are similar to those previously found in the literature. The heteroskedasticity robust F-statistics for our first stage are above 30 for $h = 0$ and increase strongly with the horizon.

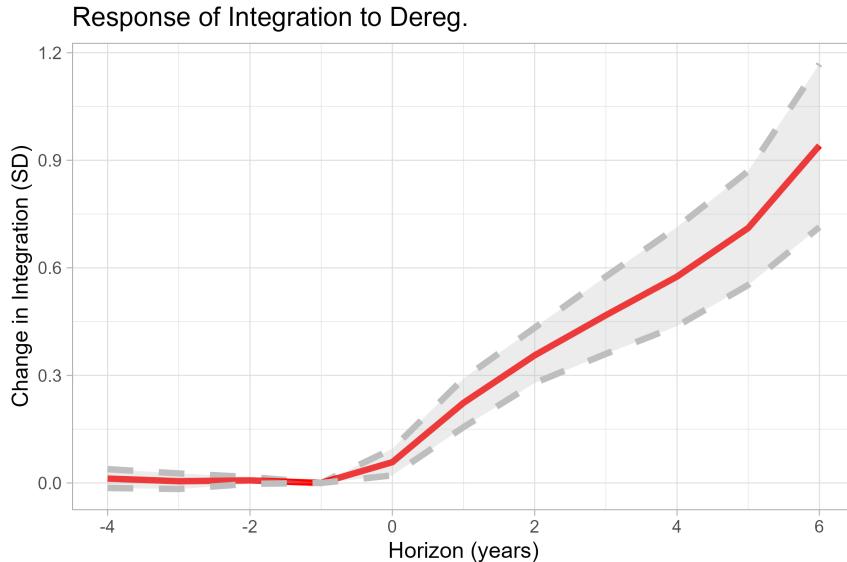


Figure 12: First stage response of standardized banking integration to banking deregulation. 90% confidence bands shown. Standard errors are dyad-robust (see text for details).

In figure 13 we plot our second stage estimates. Both consumption and output synchronization rise in response to an increase in integration, with the average response across horizons at about 2 percentage points. The estimates are uncertain and large in the short run, but stabilize and are significant at further horizons. This largely tracks with the improvement in the strength of the first stage as h increases. The magnitudes of the estimates at farther horizons are modestly larger than what a simple ratio of reduced form and first stage estimates would give⁴⁸.

⁴⁸This likely arises from differences in how the IV, reduced form, and first stage regressions weight the different treatment cohorts when calculating the aggregate effect. Following Dube et al. (2023), it is possible to reweight the IV regression to match, say, the reduced form regression's weights.

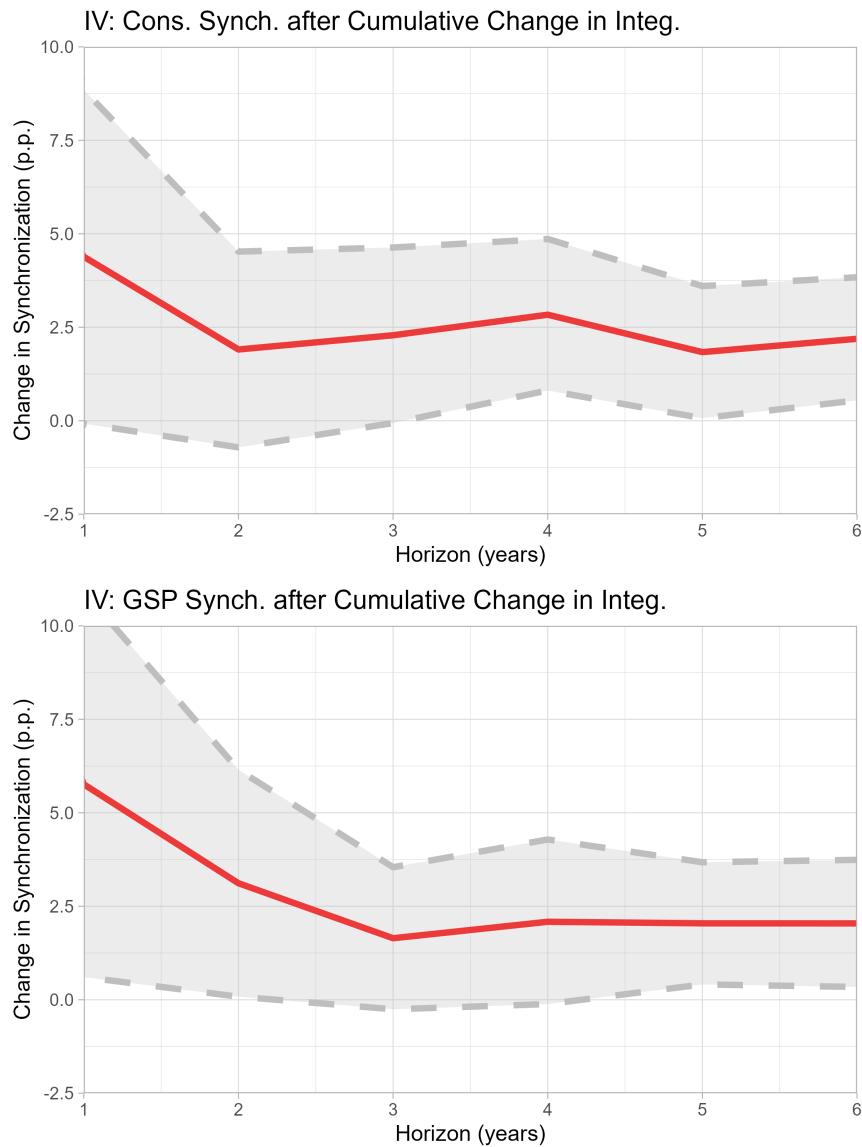


Figure 13: IV estimates of the change in synchronization to a 1 S.D. cumulative change in integration after h years. 90% confidence bands shown. Standard errors are dyad-robust (see text for details).

Note that while this IV specification is not quite a traditional impulse response it does estimate a similar, policy-relevant quantity. To estimate an impulse response, we could run $y_{i,j,t+h} - \widehat{y}_{i,j,t-1} = \beta_{IRF}^h(\widehat{\text{int}}_{i,j,t}) + \delta_t^h + e_{i,j,t}^h$. Implicitly, the β_{IRF}^h in this specification combines both a forecast integration over h periods and a forecast for $y_{i,j}$ over h periods in response to an integration “shock” at time t ⁴⁹. While the regressand is the same, the precision of the estimator depends on the strength of the time t first stage and the coefficient is normalized differently. In unreported results, we find that the β_{IRF}^h are much less precise than β_{IV}^h and suffer from weak-IV issues. Additionally, normalizing by the long-difference in integration yields a more policy-relevant quantity. Our first stage results below show that integration responds only gradually to deregulation in the short run and grows steadily over time. The size of the long-run change is likely influenced by the very low initial levels of integration. Policy interventions in different contexts will likely generate integration responses that are similarly slow-moving due to the inherent frictions in expanding banking networks; however the effects may be smaller if the initial level of integration is higher. Normalizing directly by total change in integration over a horizon can therefore give an estimate that is more “portable” to other contexts.

D.2 Additional IRFs

See figures 14 and 15.

E Additional Results for Fiscal Multipliers

E.1 Specification Choices to Address Excess Smoothness

The estimand in the consumption multiplier regressions is a standard ratio of second moments and is therefore subject to attenuation and bias due to excess smoothness. To fix ideas, consider again a simple, single-state regression on filtered consumption estimates: $\Delta\hat{c}_t = \alpha + \beta^c \Delta g_t + \epsilon_t$. As in section 4, our coefficient can be written as:

$$\begin{aligned}\beta^c &= \text{Cov}(\Delta\hat{c}_t, \Delta g_t) / V(\Delta g_t) \\ &= \underbrace{K \frac{\text{Cov}(\Delta c_t, \Delta g_t)}{\text{Var}(\Delta g_t)}}_{\text{attenuated true coefficient}} + \underbrace{\sum_{j=1}^{\infty} (1-K)^j K \frac{\text{Cov}(\Delta c_{t-j}, \Delta g_t)}{\text{Var}(\Delta g_t)}}_{\text{bias from higher order covariances}}\end{aligned}$$

Attenuation will cause underestimation of the true parameter but the bias term may be of any sign. To address this, we can control for lags of $\Delta\hat{c}_t$ in the regression. Define the estimation errors as $\nu_t = \Delta c_t - \Delta\hat{c}_t$. Then, controlling for J lags of $\Delta\hat{c}_t$ gives:

$$\begin{aligned}\beta^c &= \text{Cov}(\Delta\hat{c}_t, \Delta g_t) / V(\Delta g_t) \\ &= \underbrace{K \frac{\text{Cov}(\Delta c_t, \Delta g_t)}{\text{Var}(\Delta g_t)}}_{\text{attenuated true coefficient}} + \underbrace{\sum_{j=1}^J (1-K)^j K \frac{\text{Cov}(\nu_{t-j}, \Delta g_t)}{\text{Var}(\Delta g_t)}}_{\approx 0 \text{ due to IV, controls}} + \underbrace{\sum_{j=J+1}^{\infty} (1-K)^j K \frac{\text{Cov}(\Delta c_{t-j}, \Delta g_t)}{\text{Var}(\Delta g_t)}}_{\text{bias from further higher order covariances}}\end{aligned}$$

⁴⁹See Alloza et al. (2019) for more discussion.

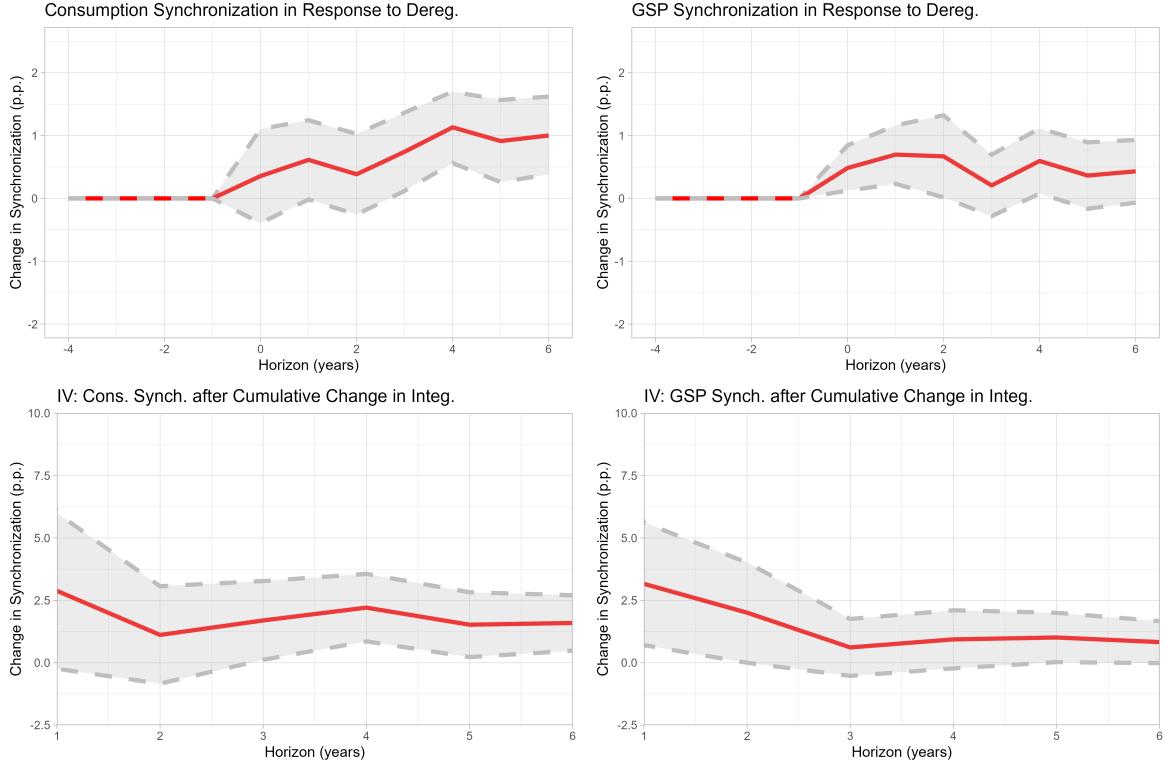


Figure 14: Reduced form (first two) and IV (second two) responses of consumption and output synchronization to banking deregulation controlling for 3 lags of the outcome. 90% confidence bands shown. Standard errors are dyad-robust (see text for details).

These estimation errors ν_{t-j} are a function of measurement errors, fundamental shocks to consumption, and prior estimation errors.⁵⁰ The main source of covariance is likely from past fundamental shocks to consumption. Since we are instrumenting Δg_t to deal with potential endogeneity between consumption and government spending, these $Cov(\nu_{t-j}, \Delta g_t)$ terms should be small in our application. Assuming covariance stationarity of Δc_t and Δg_t , the early, low j terms in the infinite sum represent a substantial fraction of the total bias. Controlling for the first J lags of Δc_t should therefore “knock out” the majority of the bias.⁵¹.

⁵⁰This can be seen by writing the Kalman filter recursions in terms of innovations to c_t and forecast errors.

⁵¹When using the two-sided or smoothed estimates for consumption, there is a symmetrical term capturing the correlation between current period δg_t and *future* values of true consumption Δc_{t+j} . One could also control for leads of consumption to address this source of bias as well.

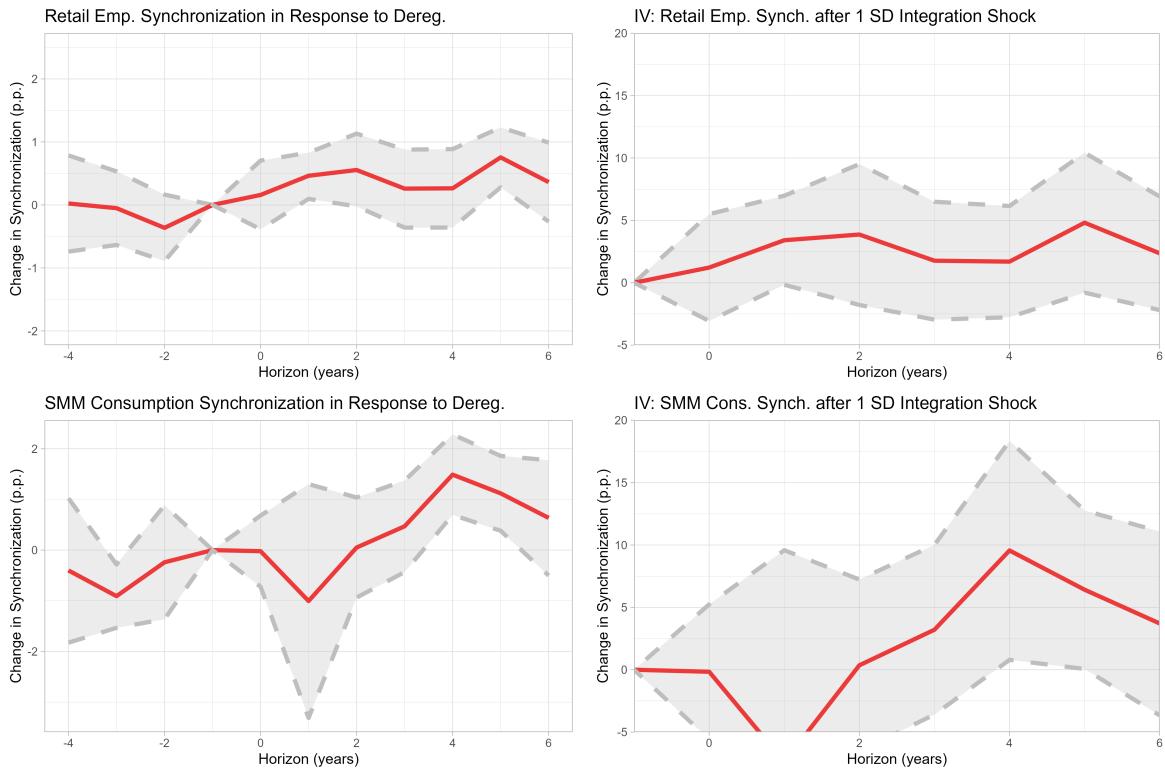


Figure 15: Reduced form (first column) and IV (second column) responses of retail employment (first row) and SMM consumption (second row). 90% confidence bands shown. Standard errors are dyad-robust (see text for details).

E.2 Short Run Fiscal Multipliers

We estimate short-run fiscal multipliers for output and consumption using the specification from Nakamura and Steinsson (2014):

$$\begin{aligned}\frac{y_{i,t} - y_{i,t-2}}{y_{i,t-2}} &= \alpha_i^y + \gamma_t^y + \beta^y \frac{g_{i,t} - g_{i,t-2}}{y_{i,t-2}} + \Gamma^y X_{i,t-1} + \varepsilon_{i,t} \\ \frac{c_{i,t} - c_{i,t-2}}{y_{i,t-2}} &= \alpha_i^c + \gamma_t^c + \beta^c \frac{g_{i,t} - g_{i,t-2}}{y_{i,t-2}} + \Gamma^c X_{i,t-1} + \eta_{i,t}^c \\ \frac{g_{i,t} - g_{i,t-2}}{y_{i,t-2}} &= \alpha_i^z + \gamma_t^z + \Pi z_{i,t} + \Gamma^z X_{i,t-1} + \nu_{i,t}\end{aligned}$$

where $y_{i,t}$ is real output per capita in state i , $c_{i,t}$ is real estimated consumption per capita, $g_{i,t}$ is real per capita military spending, X is a collection of time $t-1$ or earlier controls, and $z_{i,t}$ is either the set of interaction instruments or the single Bartik instrument. We use the national consumer price index (CPI) to convert from nominal to real quantities. Previous research has used the state-level price series constructed in Del Negro (2002), which is constructed from official city-level price indices and cost-of-living estimates from a private firm. Similar to private-sector consumption estimates, the measurement error in these series is largely unknown, so we make use of national numbers⁵². Our regressors and regresses are normalized by lagged output following Hall (2009) and constructed as second differences to address timing issues in the military spending series and anticipatory effects on output or consumption prior to spending occurring as in Nakamura and Steinsson (2014). In all regressions the standard errors are Driscoll-Kraay with two lags to account for cross-sectional dependence and serial correlation in the errors. We drop from our sample states the five states without sales taxes due to excess smoothness concerns⁵³.

Table 6: Multiplier Estimates Using NS Instrument

	Output	Cons	Output	Cons	Output	Cons
Military Shock (Bartik)	2.981 (2.821)	2.086* (0.999)	2.659 (2.740)	1.858 (0.970)	1.521 (1.342)	1.189 (0.893)
Weak-IV Robust	No	No	Yes	Yes	Yes	Yes
Lagged Outcome Controls	No	No	No	No	Yes	Yes
Observations	1575	1575	1575	1575	1485	1485
Years	1970-2006	1970-2006	1970-2006	1970-2006	1972-2006	1972-2006

Driscoll-Kraay Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In table E.2 we present estimation results using the Bartik instrument. In the first two columns we estimate the specification with no additional controls. The output response is large and imprecisely estimated, with a point estimate similar to that in NS. The consumption response is somewhat smaller and significant. In columns three and four we use the

⁵²Hazell et al. (2022) constructs state-specific price indices using BLS microdata. Unfortunately, these series' coverage across states is limited and begins in 1978.

⁵³

Andrews and Armstrong (2017) weak-IV robust estimator. Both the point estimates and the estimated standard errors are smaller. In the last two columns we control for two lags of the outcome variable as well to address persistence or other dynamics in our outcomes. These estimates are much moderated, with an output multiplier close to 1.5 and consumption multiplier close to 1.2. These last results are our preferred specification that we base our following dynamic estimates on. In the appendix, we also report results using the interactions instrument. The point estimates are somewhat smaller across the board, but the weak first stage leads to unbounded AR confidence sets. We also report OLS estimates for similar specifications and find point estimates below close to 0 for consumption and around 0.25 for output.

F Further Data Discussion

F.1 Definitions

Though this paper focuses on total retail sales, we do occasionally make use of more disaggregated data to construct our input series. Our notion of durable retail sales consists of sales by firms specializing in durable goods retailing. Under the SIC classification, this consists of establishments under the industry codes for Building Materials and Hardware Stores (SIC 51), Automotive Dealers excluding gasoline service stations (SIC 55 excl. 554), and Furniture and Appliance stores (SIC 57). Under the NAICS system, the relevant industry categories are Automotive (NAICS 441), Furniture (NAICS 442), Electronics (NAICS 443), and Buildings Materials (NAICS 444). The remaining types of sales are classified as non-durable, and consist chiefly of sales by food stores, drug stores, apparel stores, gas stations, eating establishments, miscellaneous stores, and department stores.

F.2 Tax Collections by Industry

The data we collect are at the industry by state level, although the level of industry detail varies by state. In most cases, industry sales tax receipts are available at the 2 digit SIC level (e.g.: SIC 54: Food Stores, SIC 56: Apparel Stores, etc.), although some states offer detail at the 3 digit level, and others use non-standard classification schemes. This information is published by states in an idiosyncratic way, under a variety of titles, including Washington State's *Quarterly Business Review*, the Mississippi *Tax Comission Service Bulletin*, and Colorado's *Department of Revenue: Annual Report*. Most of the series constructed from these reports begin in the 1970s, and run through the early 2000s, although the precise span of each series also varies on a state basis.⁵⁴ In the appendix we discuss the reports used for each state, many of which we were only able to locate with the help of state librarians.

We find it useful to distinguish between three types of receipts reports:

1. *Gross Sales*: Gross sales entail all receipts reported by an establishments, including those which are not subject to the sales tax. This measure is ideal in that it would not

⁵⁴We currently end many of these series in the early 2000s as a result of two complications: first, many states made the categorization transition from SIC to NAICS during this period, introducing a change in the series definition. Second, the rise of e-commerce causes some of the retail trade sub-categories to exhibit unstable variation.

vary with legal changes in the coverage of the tax base. Unfortunately, most states do not reliably report gross sales.

2. *Taxable Sales*: Taxable sales entail all receipts reported by an establishments which are subject to the sales tax. In most states, this differs from gross sales by the value of food for home consumption, as well as gasoline. Several states report a measure of taxable sales, which have the advantage of not requiring any adjustment by sales tax rates.
3. *Tax Revenues*: These are simply revenues collected by the state (sometimes net of refunds). These fluctuate as a result of both base changes and rate changes.

As discussed previously, sales taxes have 3 main limitations as measures of retail sales:

- 1.) Inclusion of intermediates and goods not for households
- 2.) Shifts in tax bases and rates
- 3.) Exclusion of food items. In the case where we have gross sales information by industry, we overcome all three of these challenges: 1.) the industry coding allows us to exclude non retail trade establishments, 2.) the gross sales values is invariant to base shifts and 3.) even food items are reported under the gross sales heading. The red line plotted in Figure (1) is constructed from gross sales figures that Washington State classifies as retail trade establishments.

In practice, we usually have information on either taxable sales or tax revenues, which are still useful for overcoming the first two limitations. We can overcome the first challenge whenever our industry break-down allows us to exclude sales tax revenues from non retail trade establishments. And although both taxable sales and revenues will vary with the legal form of the tax base, information on sub-categories of retail trade (e.g., apparel, or gas service stations) allow us to calculate adjusted growth rates that ignore the establishments experiencing a legal adjustment. In practice, many of the revisions to the tax base are changes in the treatment of the service sector, which we typically exclude wherever we have access to industry breakdowns.

An additional advantage of having industry breakdowns for sales tax revenues, is that they contain information on the relative sales of durable and non-durable establishments. Along with the information on non-durable sales from the geographic MRTS, this allows us form estimates of non-durable sales growth for a sub sample of states.

F.3 Retail Census

We use the retail census to construct a panel of 5 year growth rates in retail spending for each state over the period of 1967 - 2017. In this section we discuss the details of how these growth rates are constructed. We focus on 5 year growth rates because, since 1967, the census has been conducted in years ending in 2 and 7 (although we also have a growth rate for the 4 year change corresponding to the 1963 - 1967 period). Since 1992, the census data have been posted on the Census Bureau's website. For earlier years, scans of the census tables can be found on archive.org. Under the SIC system prevailing before 1997, the census covers the industry codes 51 - 59. As discussed in the main text, this includes "eating and drinking places (SIC 58)" which under the contemporary NAICS system is classified as a

service. After the NAICS system is introduced in 1997, we use the industry codes 44-45, in addition to the food service category (NAICS 722).

Our growth rates are constructed from the total sales volumes reported for all retail trade establishments, with two caveats. First, we subtract the sales of "Nonstore retailers (SIC 596)" that specialize in cross-border sales, before calculating the growth rate. Second, we use figures for "Establishments with Payroll." Before 1982, the census reported sales for all establishments (including those which did not report payroll tax). Unfortunately, the data for non-payroll establishments seems to have required cooperation from the IRS, which was not forthcoming after 1982. We choose to use the payroll establishment figures for the whole sample. For reference, in 1972, payroll establishments accounted for a median (across states) of 96% of total retail sales, so relatively little is lost from the exclusion of non-payroll establishments. The correlation (across states) between the 1967 to 1972 sales growth rates for 'all retail establishments' and 'payroll retail establishments' is over 0.99.

As discussed in the text, there were adjustments to the definition of retail sales made between some of the censuses. We now discuss details of how each growth rate was computed:

- 1963 - 1967: To compute this 4 year change, we use Table 1, parts A and B, printed in the 1967 retail census books. There do not appear to have been major revisions to the retail definition over this period.
- 1967 - 1972: This change is computed from the 1972 census books, table 7, parts A and B. Over this period, several establishments specializing in electrical and plumbing supplies were re-classified from the retail to the wholesale sector. Table 7 presents the 1967 and 1972 records tabulated under a unified definition, allowing us to compute a growth rate unaffected by this classification change.
- 1972 - 1977: The core source for this change is Table 2 in the 1977 retail census books, which presents the 1972 and 1977 value under the same industrial classification.

Unfortunately, there was a change to the meaning of "sales" over this period: in 1972, respondents were asked to include sales taxes and finance charges in their reported sales, but were asked to exclude these items in 1977. This makes the 5 year growth rates for this period a lower bound on the true growth rate, and the tightness of this bound will depend on each state's degree of sales taxes and financing charges.

To mitigate this problem, we adjust the 1972 census figures by an estimate of the degree of sales taxes and financing charges that were reported in each state. The adjustment happens in four steps:

1. First, we estimate the degree of financing charges in each state. We begin by disaggregating each state's retail sales into 10 categories at the 2 digit SIC level (e.g., Apparel Stores (SIC 56), Furniture Stores (SIC 57)).⁵⁵ For each of these industries, we have an estimate of the national percentage of sales which constitutes

⁵⁵Technically a few of our disaggregated categories are at the 3 digit level: Gasoline Service Stations (SIC 554) and Drug Stores (SIC 591). These categories were chosen based on the organization of the Census booklets.

financing charges: these estimates are taken from Table 4 of the Census Bureau's *Current Business Reports: 1978 Retail Trade, Annual Sales, Year-End Inventories and Accounts Receivable by Kind of Retail Store*.⁵⁶ For example, general merchandise establishments (primarily department stores) are estimated by the Census Bureau to have financing charges of 2.2%, while the retail trade sector as a whole is estimated to have financing charges of only 0.5%. To form our estimate of each state's financing charges, we simply multiply each industry's sales by its national financing charge ratio, and take the sum. States with a higher percentage of retail sales at the department store and automotive establishments have higher estimated financing costs, relative to their total sales.

2. Second, we estimate the degree of sales taxes reported for each state. We do this using two distinct methods. The first, which is entirely analogous to our method for estimating financing charges, is to weight each state's 2 digit SIC retail sales by an estimate of the national sales tax percentage. This time the national percentages are taken from the Census Bureau's *1976 Retail Trade* report, which is similar to the 1978 report cited previously. This report indicates the retail sector had a sales tax rate of 3.1%, with food stores facing a rate of only 1.9%.

The second method is to directly use information on sales taxes which prevailed in each state in 1972. In this case, we also perform some sectoral disaggregation where appropriate: we assume that only 25% of food stores sales are taxed in states known to exempt groceries in 1972 (we use special, reduced tax rates on groceries for Louisiana and the District of Columbia). This method will still feature some degree of error, as there may be local level sales taxes for which we have no systematic information, as well as small and idiosyncratic sales tax exemptions which cannot be exactly measured for each sector.

Since each method features some error, our total estimate of sales taxes reported in the 1972 census is simply the average of the two. We think of this as a "crude Bayesian" adjustment. In the absence of any information about any state's specific tax rates, we would rely on the first method. If we had perfect information on the tax burdens by state, we would use only the second. With imperfect information, we choose to blend the two estimates.

3. The third step is to re-scale our estimates of financing and tax charges to match an aggregate value reported by the Census Bureau. In the appendix to the 1977 retail census books, the Census estimates that in 1972 there were 10 billion dollars of financing charges and sales tax reported by surveyed retail establishments. Since there were 457.4 billion dollars of total retail trade sales for the entire nation, the total amount of sales we wish to subtract from the 1972 census should sum to $\frac{10}{475.4} \approx 2.19\%$ of the national total.

The Census also states, in its discussion of the 1972 tax and financing charges, that there was considerable under-reporting of these charges during the 1972 census. This is consistent with what we find, which is that our estimate of the aggregate charges (formed by summing of estimate of the charges for each state) is

⁵⁶Currently this report is hosted on the site Hathi Trust, availalbe at this URL <https://catalog.hathitrust.org/Record/002598329>

larger than the 2.19% figure which we target. Our third step is to multiply each state's charge estimate by the ratio of the census's internal charge estimate to our national charge estimate. This guarantees that the sum of all the charges we subtract for each state matches the 2.19% value estimated by the Census Bureau.

4. Finally, we subtract the charge estimate for each state from the sales for that state. This gives us an adjusted estimate of 1972 retail sales, which conceptually aligns with the definition of sales used in 1977 (i.e., a definition that excludes tax and financing charges).

Thus, for every state, our "adjusted" 1972 census series features lower sales than the headline 1972 census. If we had adjusted every state down by the same percent, each state would have 2.19% lower sales than the headline figure. Instead, our procedure makes larger downward adjustments for states with higher sales tax burdens and more activities in sectors with high financing and tax charges. For example, we adjust Delaware's sales down by only 1.35% (the state has no sales tax), while we reduce Connecticut's reported sales by 2.64% (the state had a relatively high sales tax burden at the time). One might argue that our adjustment for Delaware should be closer to 0.5% (without sales taxes, there would be only financing charges), but the "Bayesian" averaging in our procedure "shrinks" each adjustment toward the national adjustment.

Ultimately, our 1972 -1977 growth rate uses the 1977 sales from Table 2, along with the 1972 values, after being adjusted using the above process.

- 1977 - 1982: This change is computed from the 1982 census books, Table 2. Over this period the Census Bureau changed its approach for classifying leased department stores. The new system seems to have slightly reduced total retail sales for the nation, as well as changed sectoral numbers within states. Fortunately, the figures in Table 2 offer growth rates with an internally consistent classification.
- 1982 - 1987: This change is computed from the 1987 census books, Table 3. Over this period, the SIC system changed its guidelines for classifying department stores and some electronics establishments. Table 3 presents the 1982 and 1987 figures under an internally consistent classification.
- 1987 - 1992: We compute this change from Table 3 in the 1992 census books. In this case, the figures for 1987 seem to exactly match the headline figures from the 1987 books.
- 1992 - 1997: In 1997, the Census Bureau began classifying establishment using the NAICS system. This means the headline figures published for 1997 are not comparable to the sales totals for 1992. Fortunately, the Census Bureau published an auxiliary document which re-tabulates the 1997 data using SIC classifications at the state level.⁵⁷

⁵⁷This document is named 1997 Economic Census Core Business Statistics Series: Comparative Statistics. It is can currently be found at this link : <https://www.census.gov/library/publications/1997/econ/census/core-business-statistics-series.html>

We compute our growth rates using the 1992 SIC totals and the re-tabulated 1997 SIC values released in the auxiliary tables.

Unfortunately, there is one complication with the auxiliary tables mentioned above: for 5 states, the total retail trade sales are censored to prevent identification of the sales of individual companies.⁵⁸ For example, in California, total retail trade sales are censored, along with sales amounts for department stores and home furnishings stores. Fortunately, sales amounts are still published for the other major retail trade sectors (apparel, automotive, etc.).

To deal with this issue, we perform imputations at the narrowest industry levels possible (so in the case of California, we work to impute the sales for department stores and home furnishing stores, allowing us to calculate the sum for the retail trade sector). The procedure of our imputations roughly works as follows:

1. Identify a 3 digit SIC industry that requires imputation. For example, in California, we need to impute sales for Home Furnishing and Furniture Stores (SIC 571). Further identify all 4 digit SIC sub-industries which aggregate to the 3 digit industry. In this case, SIC 571 is an aggregate of SIC 5712 and SIC 5713.
2. Using the Census Bureau's *1997 Bridge Between SIC and NAICS*, identify all 8 digit NAICS codes which map into each 4 digit SIC code.⁵⁹ For example, in the bridge, SIC Code 5712 is the sum of the following seven 8 digit NAICS Codes: 33711030, 33712130, 33712220, 44211010, 44211020, 44211030, 44211041. Meanwhile, SIC Code 5713 is the sum of a single 8 digit NAICS Code, 44221010. This means that the 3 digit value we wish to impute, SIC 571, is the sum of eight different 8-digit NAICS codes.
3. We then impute values for these 8-digit NAICS codes. Since 8-digit sales are never published at the geographic level, we first find the most narrow geographic sales figure available. For example, the 1997 NAICS Tables for California contains sales information for the 6-digit NAICS 442110. This 6-digit code is a sum containing the sales of several of the 8-digit codes we want to estimate (4 of 8-digit codes we want, to be specific). Our goal is now to determine what fraction of those 6-digit sales were in the 8-digit bins we care about.
4. To estimate these fractions, we rely on national data. The bridge table presents 6 and 8 digit sales information at the national level. For the nation, the sales of the 8 digit NAICS 44211010 were 75.2% of the sales of the 6 digit NAICS 442110. Our assumption is that for California, this fraction was the same, and so our estimate of the 8-digit NAICS value is the 6-digit value times 75.2%.
5. We sum all of our 8-digit NAICS sales estimates to recover an estimate of the relevant 3-digit SIC sales amount.

⁵⁸Sales totals are missing for California, Washington, Wyoming, South Dakota, and the District of Columbia. The Census appears to have censored an unusually large number of data cells in these tables, as the publication of comprehensive NAICS and SIC tables considerably increased the information available to identify establishments in narrow industry categories. Ther eis a brief discuss in the linked document, p. 12.

⁵⁹The bridge pdf can be found at this link <https://www2.census.gov/library/publications/economic-census/1997/core-business-statistics-series/97x-cs3.pdf>

The exact details of the procedure vary somewhat from state-industry to state-industry. In some cases we can be precise about the sum of two sub-industries sales, even when the exact sales of each industry are unknown. In a handful of cases, stronger assumptions are needed to impute a NAICS sales value.

After completing imputations for 5 states, we calculate our final 5 year SIC retail sales growth rates.

- 1997 - 2002, 2002 - 2007, 2007 - 2012, 2012 - 2017: Each of these 5 year changes is computed from the Economic Census NAICS Tables on the Census Bureau's website. There were no important changes to the NAICS retail definition over this period. In addition to including sales for the 44-45 NAICS Codes, we include food service sales (NAICS 722). We continue to subtract the sales of non-Store retailers, which become large as e-commerce grows through the 2000s.

F.4 MRTS

Data construction:

A contribution of this paper is to extend the *Monthly Retail Trade Survey* data back to 1968 for large areas. Previous work had used data for the 1978 - 1996 sample. More recently researchers appear to have lost the data for 1978 - 1985; the work (Mian et al. (2020)) uses only data from 1986 on, presumably because this is the only data currently published on the Census Bureau's website.

The key challenge in constructing these series has been that the Census Bureau occasionally revises its retail definitions and sampling practices. These revisions cause discrete jumps in the time series which do not reflect economic activity. Our approach to this problem is to construct 6 internally consistent vintages of the data. Fortunately, each of these vintages overlaps with an adjacent vintage, which allows us to compare how similar the growth rates of each vintage are. Reassuringly, where vintages overlap, the growth rates tend to have correlations over 0.99, even where the levels are different.⁶⁰ Our final series simply uses the percent change from the latest available vintage at each point in time, with a level normalized to the last vintage.

We detail the sources for these six vintages. All data are collected from publications in the Census Bureau's *Current Business Reports Series*, but we offer further specifics below:

1. **Final Vintage (Vintage 1), Jan 1982 - Dec 1996:** The data for the period of Jan 1987 to December 1996 are sourced from the publication *Current business reports. Annual benchmark report for retail trade. BR/96-RV*.⁶¹ Data for the period of Jan 1982 - December 1986 are sourced from editions of the the report *Current Business Reports. Revised Monthly Retail Sales and Inventories*.⁶² The final report in the

⁶⁰Thus, it appears that the Census Bureau revisions tend to be mostly be some kind of re-scaling, which has no effect on growth rates. The Census Bureau suggests that some of its revisions are intended to benchmark the monthly survey results against results from the Annual Retail Trade Survey, and this may result in re-scaling.

⁶¹This document is currently available on HathiTrust at this link <https://catalog.hathitrust.org/Record/003192678>

⁶²These reports are available on HathiTrust at this link : <https://catalog.hathitrust.org/Record/003003480>

Revised Monthly Retail Sales Series (edition BR/94-RV, Jan 1985 Through Dec 1994) agrees exactly with the *Annual Benchmark Report* for the period of Jan 1987 - Dec 1992; there are small discrepancies over 1993 and 1994, where we defer to the *Annual Benchmark Report*, which was published later. For the years of 1986 - 1996, this produces a series that exactly matches the tables published on the Census Bureau's website.⁶³

The data for Jan 1982 - Dec 1984 are taken from edition BR/91-RV (Jan 1982 Through Dec 1991) of the *Revised Monthly Retail Sales Series*. The 1985 and 1986 tables in this series agree exactly with the later report BR/94-RV (which in turn agree with the Census website), reassuring us that the definition of retail trade used across these documents is not subject to major changes in scope or re-indexing.

2. **Vintage 2, Jan 1978 - Dec 1985:** In the late 1980s, the Census Bureau revised the retail sales series in a way that shifted their levels. Since 1982 was the earliest year for which it issued the revised series, the data for 1981 and earlier are simply discontinuous with the final data series. What we call Vintage 2 is an internally consistent time series for the period of Jan 1978 to Dec 1985, released before the late 1980s revision. The data for Vintage 2 were collected from early edition of the *Revised Monthly Retail Sales Series* discussed above. The specific editions are BR 89-R (Jan 1980 Through Dec 1989) and BR-13-87S (Jan 1978 Through Dec 1987).

There are several years of overlap between Vintages 1 and 2 in the mid-1980s. This allows us to compare whether the monthly dynamics of the series are different. Fortunately, they are extremely similar, as the monthly percent changes have a correlation of over 99%. Thus, our approach to building a 1978 - 1996 time series is simply to use the percent changes from Vintage 2 to extend Vintage 1 backwards in time.

It is unclear whether previous users of the 1978 - 1996 sample (Hess and Shin, Del Negro) were aware of the revision problem.

3. **Vintage 3, Jan 1976 - Dec 1978:** The earliest date for which there is Vintage 2 data is Jan 1978. However, for 15 large states (and the 9 Geographic Divisions), we can use auxiliary *Current Business Reports* to extend the time series further. Again, these reports were published before level revisions were made to Vintage 2, so we are careful not to splice them together in levels.

Our source for this vintage is the series *Current Business Reports. Monthly Retail Trade. Sales and Accounts Receivable*. The data for Jan 1978 to Dec 1978 are sourced from the December 1978 edition (BR-78-12).⁶⁴ This was the final edition published before a sequence of level revisions went into effect. The data for Jan 1977 to Dec 1977 are taken from the Jan 1978 volume (BR-78-01), and the data for Jan 1976 to Dec 1976 are from the September 1977 report (BR-77-09).

The September 1977 report is very explicit that there was a level revision at this time (indeed, the revision was related to the exclusion of sales taxes from sales figures, discussed in the previous appendix section). Fortunately, this report republished the 1976

⁶³Currently hosted at this link <https://www.census.gov/retail/mrts/mrtshist.html>

⁶⁴Some of these volumes can be found on HathiTrust <https://catalog.hathitrust.org/Record/002449621>

data under the new conventions adopted in 1977. We carefully read every report between September 1977 and December 1978 to ensure there were no other level revisions over this period.

4. **Vintage 4, Jan 1970 - Dec 1976:** According to the September 1971 report (BR-71-09), there was a large revision enacted in 1971. This report re-tabulates the 1970 data using the new procedures, and thus serves as our Jan 1970 - Dec 1970 source for this vintage. We carefully read all the monthly reports over the September 1971 to September 1977 period to ensure there are no other major revisions. For each year between 1971 and 1976, the source for this vintage is the report titled with January of the following year.

As an additional data quality check, we compare the values from these monthly reports to annual sums reported in the series *Current Business Reports: Annual Sales and Purchases, Year-End Inventories, and Accounts Receivable, by Kind of Business*. (the 1973 volume is BR-73-13). These reports were published some months after the monthly versions, giving the Census Bureau time to publish any additional updates to the data. We find the series match across the annual and monthly reports, with the exception of a small typo for New York's 1971 non-durable sales.

5. **Vintage 5, Sep 1967 - Dec 1970:** The September 1968 report (BR-68-09) explains that an important sample revision was implemented in late 1968. The data for late 1967 and 1968 were taken from this report, along with the January 1969 report. Data for the years of 1969 and 1970 were taken from the Jan 1970 and Jan 1971 reports, and double-checked against the *Annual Sales and Purchases* volumes. These sources agreed, except for a small discrepancy related to a revision of California's 1969 data.
6. **Vintage 6, Jan 1966 - Sep 1967:** The very earliest geographic retail sales report was published for April 1963, with 12 months of information for 9 large states.⁶⁵ After reading the remaining reports for the 1960s, we verified there is no need to define additional vintages of the data. However, some of the monthly fluctuations in the data over the 1963 - 1966 period are extreme to the point that we do not believe the Census was using the same sampling or statistical procedures that it employs for the majority of the survey's life. For that reason, we drop values from before 1967.

Geographic Divisions Definition:

Geographic divisions: New England: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut; Mid Atlantic: New York, New Jersey, Pennsylvania; West North Central: Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas; East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin; South Atlantic: Delaware, Maryland, DC, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida; East south Central: Kentucky, Tennessee, Alabama, Mississippi; West South Central: Arkansas, Louisiana, Oklahoma, Texas; The West: Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada; Pacific: Washington, Oregon, California, Alaska, Hawaii

⁶⁵The report can be found here: <https://babel.hathitrust.org/cgi/pt?id=pst.000072746134&seq=1>

G Sampling with Sequential Monte Carlo

We are interested in conducting inference on functions $h(\theta)$, such as the model parameters themselves or latent state-level consumption. This requires characterizing the model's posterior $p(\theta|y)$. Let $y = \{y_t\}$ be the dataset, $p(\theta)$ the prior, and $p(y|\theta)$ the likelihood from the extended Kalman filter. Then,

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

For convenience, we follow Herbst and Schorfheide (2016) and define $\pi(\theta) = p(\theta|y)$, $f(\theta) = p(y|\theta)p(\theta)$ and $Z = p(y) = \int p(y|\theta)p(\theta)d\theta$ so that $\pi(\theta) = f(\theta)/Z$.

A standard approach to this problem is to draw samples $\{\theta^i\}$ from the posterior $\pi(\theta)$ using random walk Metropolis Hastings (RWMH) in order to estimate a Monte Carlo average $\bar{h}(\{\theta^i\}) = \sum_{i=1}^N h(\theta^i)$. As is well known, this sampling routine is quite inefficient when $p(\theta|y)$ is high-dimensional or multimodal. Common issues include a tendency to "get stuck" in local modes and high autocorrelation of draws. These can lead to an inaccurate approximation of the posterior and have been highlighted as important obstacles to estimating medium-sized DSGE models using RWMH. Our model has an order of magnitude more parameters than those typically have and is therefore a poor candidate for RWMH.

We instead make use of sequential Monte Carlo (SMC), an alternative sampling algorithm that has found use in estimating DSGE models. SMC techniques were originally developed for nonlinear particle filtering but have found use in estimating model parameters as well. SMC can be viewed as an augmented form of importance sampling, where draws from our desired distribution $\pi(\theta)$ are created by reweighting draws from an easier-to-sample proposal distribution $g(\theta)$. Formally, this starts with the identity

$$E_\pi[h(\theta)] = \int h(\theta)\pi(\theta)d\theta = \frac{1}{Z} \int h(\theta)w(\theta)g(\theta)d\theta, \text{ where } w(\theta) = \frac{f(\theta)}{g(\theta)}$$

where $w(\theta)$ is an importance weight that transforms draws from $g(\theta)$ into draws from $\pi(\theta) = f(\theta)/Z$. Given a size N iid sample $\{\theta^i\} \sim g(\theta)$ we can therefore calculate normalized weights for each draw as $W^i = w(\theta^i)/\sum_{j=1}^N w(\theta^j)$. The combined set $\{\theta^i, W^i\}$ represents our approximation of $\pi(\theta)$ and are typically called "particles" in line with their origin from particle filtering. The corresponding Monte Carlo average approximation is:

$$\bar{h}(\{\theta^i, W^i\}) = \sum_{i=1}^N h(\theta^i)W^i \xrightarrow[N \rightarrow \infty]{a.s.} E_\pi[h(\theta)]$$

Convergence to the true expectation occurs under mild conditions on h and is discussed further in Geweke (1989) and Herbst and Schorfheide (2015).

If the prior and posterior differ substantially then this simple procedure may yield few useful draws, i.e. ones from high-probability regions of the posterior, and therefore have large Monte Carlo errors. This will manifest in a large variance of the weights W^i . The core insight of SMC is that even if taking a single importance sampling step from the prior to

the posterior fails, a series of smaller steps can still succeed. In practice, this mean doing importance sampling iteratively on a sequence of bridging distributions that begins with the prior and ends with the posterior. Following the treatment in Herbst and Schorfheide (2016) we index these bridge distributions with n :

$$\pi_n(\theta) = \frac{f_n(\theta)}{Z_n} = \frac{[p(y|\theta)]^{\phi_n} p(\theta)}{\int [p(y|\theta)]^{\phi_n} p(\theta) d\theta}, \quad n = 1, \dots N_\phi$$

where ϕ_n is an increasing sequence of “tempering” parameters with $\phi_0 = 0$, $\phi_{N_\phi} = 1$. This sequence of π_n transforms gradually from the prior into the likelihood as $\phi \rightarrow 1$.

The basic SMC algorithm proceeds as follows. For each stage n , incoming particles $\{\theta_{n-1}^i, W_{n-1}^i\}$ are first *corrected* by reweighting them to reflect the change between π_{n-1} and π_n . If the distribution of particles has degraded sufficiently so that the variance of new weights W_n^i is large, particles are *selected*. In this (optional) step, we resample the particles according to their weights W_n^i to remove poor performers. While this introduces noise into the approximation, it helps to keep the number of effective draws large. Lastly, particles are *mutated* using Metropolis Hasting steps to transform them from θ_{n-1}^i to θ_n^i . These MH steps are done using π_n as the stationary distribution. This adapts each particle to the current bridge distribution and further improves the accuracy of the approximation. This sequence is initialized with draws from the prior and uniform weights: $\{\theta_0^i, 1\}, \theta_0^i \sim p(\theta)$. The specific SMC implementation we use from Cai et al. (2021) includes many improvements on the basic scheme outlined above to improve performance for difficult posteriors. These are described further in the referred to paper and generally involve adaptive features that ensure the particle distribution does not decay precipitously step-to-step and that the MH steps during mutation succeed reasonably often.

Previous work on related models in economics and on particle filtering methods in statistics support the use of SMC in this context. First, the model presented here has antecedents in work by Aruoba et al. (2016) and others that aims to “filter out” measurement error in macroeconomic time series by using observations from multiple noisy indicators. These first authors have shown that the combination of classical measurement errors with an otherwise identified ARMA model preserves identification; our setting is analogous as the underlying unobserved components model is identified (Nelson et al. 03). Second, recent research on adapting particle filters for use in high-dimensional settings suggests that our SMC implementation should perform well here. Particle filters and SMC both attempt to approximate some unknown posterior distribution using a series of importance sampling steps and therefore have similar theoretical properties. Both practitioners and theoretical researchers have noted that standard particle filtering methods, which use only the *mutation* and *correction* steps from above, often fail in high dimensions⁶⁶. Bengtsson et al. (2008) show this formally, proving that as the number of dimensions increases the distribution of weights converges to a point mass. In essence, a particle filter in high dimensions will generate only a single draw from the posterior rather than N draws. Thankfully, subsequent work by Rebeschini and Van Handel (15) show that by *resampling* particles and making basic modifications to the mutation step the weight distribution can be stabilized given a fixed N and arbitrarily large number of dimensions. These modifications were already incorporated in the SMC

⁶⁶For SMC, a higher dimension corresponds to a longer parameter vector θ .

implementation we use and as such the results of Rebeschini and Van Handel should carry over to our setting as well.