U.S. State-Level Business Cycles Since the Civil War

Joseph Hoon Chang Liu Karsten Müller Zhongxi Zheng

National University of Singapore

August 26, 2025

EEA 2025

Motivation

- US states provide a rich environment to
  - study regional heterogeneity
  - identify factors underlying macro dynamics (e.g., business cycles and long-run growth)
  - identify the transmission of economic policies (e.g., fiscal and monetary policy)
- However, many state-level economic indicators have limited availability
  - BEA reports state-level GDP starting from 1963
  - Many short time series available for other state-level macro indicators such as consumption
- Stark contrast with work of economic historians on specific time periods or aggregate series
  - Large literature on the Great Depression using *regional* data
  - Recent work using *aggregate* series—e.g., Payne et al. (2025); Bouscasse, Nakamura and Steinsson (2025)

This paper

- Construct a novel, long-run dataset on US state-level economic indicators
  - 1863–2021
  - 60 variables from various sources, many newly digitized
  - Harmonization of time series, due to varying definitions, data sources, and availability
- Use a mixed-frequency dynamic factor model to estimate an annual *index* of state-level economic activity
  - Highly correlated with state-level GDP (post-1963)
  - Substantial heterogeneity across states in terms of business cycle volatility and their underlying drivers
  - Business cycles have become more synchronized since WWII
- Construct indicators for state-level recessions

Data Construction

Data sources

- Starting point:
  - Statistical Abstract of the United States, available annually since 1878
  - Census reports, available decennially since 1790
- 111 additional sources of historical data:
  - We manually collect and in many cases digitize these data
  - Example alternative sources include:
    - Government reports and company surveys
    - Archives, books, almanacs, etc.
    - Previous studies in the economic history literature (e.g. Sylla et al., 1993)

Data sources

COMMERCIAL FAILURES IN 1895.

STATES.	COMMERCIAL FAILURES. TOTAL, 1895. No.	Assets.	Liabilities.	TOTAL, 1894. No.	Liabilities.	CLASSIFIED FAILURES, 1895. MANUFACTURING. No.	Liabilities.	TRADING. No.	Liabilities.	OTHER COM'L. No.	Liabilities.	BANKING. No.	Liabilities.
Maine	188	$496,156	$1,257,858	251	$2,449,210	35	$482,390	132	$774,158	1	$1,500	1	$80,000
New Hampshire	58	213,721	405,644	47	326,646	11	75,800	47	329,844	..	.....	..	.....
Vermont	36	64,400	145,300	33	315,096	6	34,000	30	111,300	..	.....	..	.....
Massachusetts	567	4,342,003	10,942,638	836	16,467,631	288	5,376,080	270	4,849,891	9	716,667	1	45,000
Connecticut	254	1,786,236	2,442,980	253	1,821,143	73	1,704,110	177	728,870	4	10,000	2	530,000
Rhode Island	202	573,925	3,771,397	187	1,460,566	46	2,866,511	137	855,486	19	49,400	1	1,166,526
New England	1,305	$7,476,441	$18,965,817	1,607	$22,860,292	459	$10,538,701	813	$7,649,549	33	$777,567	5	$1,817,526
" 1894	1,607	9,889,439	22,860,292			453	10,499,011	1,140	12,014,956	15	346,325	1	125,000
New York	1,940	$23,033,614	$45,225,534	1,976	$36,858,225	569	$25,985,159	1,344	$17,616,587	26	$1,623,788	4	$2,647,179
New Jersey	182	2,655,203	3,612,488	213	3,351,765	62	2,462,601	113	858,789	7	291,098	..	.....
Pennsylvania	1,349	7,494,071	11,739,947	1,433	15,665,058	368	4,566,682	975	7,121,065	6	52,200	6	745,454
Middle	3,471	$33,182,888	$60,577,969	3,621	$55,895,049	999	$33,014,442	2,432	$25,596,441	49	$1,967,086	10	$3,392,633
" 1894	3,621	31,377,202	55,895,049			1,150	28,415,912	2,399	26,415,795	72	3,065,402	13	7,452,724
Maryland	299	$1,825,784	$3,279,124	239	$2,971,319	86	$1,467,362	205	$1,671,706	8	$140,056	..	.....
Delaware	68	194,250	460,100	61	936,770	19	170,500	38	289,600	..	.....	..	.....
Dist. Columbia	58	367,386	713,609	52	834,455	7	92,460	50	585,824	1	35,325	1	410,884
Virginia	305	1,733,394	2,929,414	267	2,626,042	27	446,009	275	2,151,414	3	332,000	2	1,198,402
West Virginia	69	402,572	691,324	100	532,279	17	263,333	51	426,791	1	1,000	..	.....
North Carolina	103	799,317	1,231,747	139	1,901,810	5	72,000	98	1,159,747	..	.....	..	.....
South Carolina	102	1,097,763	1,263,903	97	2,121,815	8	358,453	94	905,450	..	.....	..	.....
Florida	136	1,933,400	1,741,550	46	980,650	4	31,000	131	1,704,650	1	3,500	3	450,000
Georgia	214	2,093,006	3,949,383	347	4,756,118	21	1,191,750	188	1,773,833	5	83,800	2	365,000
Alabama	149	757,000	1,339,250	190	2,944,309	11	250,500	129	1,089,750	..	.....	..	.....
Mississippi	115	829,720	935,340	171	1,357,699	1	4,000	113	930,340	1	1,000	..	.....
Louisiana	199	2,743,793	2,876,981	230	1,807,799	17	291,348	181	2,570,733	1	12,900	2	296,952
Texas	273	2,416,622	2,646,632	347	3,141,249	23	633,231	247	1,812,569	3	201,770	1	107,800
Kentucky	274	2,310,004	3,042,045	339	3,407,820	38	1,859,961	235	1,181,384	1	700	..	.....
South	2,355	$19,703,921	$26,189,502	2,635	$31,230,544	275	$7,136,160	2,055	$18,233,191	25	$811,151	11	$2,374,268
" 1894	2,635	25,454,259	31,230,544			203	9,860,311	2,304	19,450,990	28	1,919,193	12	952,754

STATE DEBTS, VALUATION, AND TAXES.

THE following statistics of the finances of the thirty-eight States in the Union have been derived in most cases from the officers of the States themselves.

STATES.	Date of Statement.	AMOUNT OF STATE DEBT. Funded.	Unfunded.	Amount Raised by Taxation Last Year.	AMOUNT OF TAXABLE PROPERTY AS ASSESSED Real.	Personal.	State Tax Per Cent on $100.
		$	$	$	$	$	Cts.
Alabama	1876.	14,061,670	.....	.....	83,851,252	76,200,000	75
Arkansas	Sept. 30, 1877.	4,153,035	13,967,012	457,450	61,960,452	32,692,425	60
California	June 30, 1877.	3,411,000	.....	4,105,884	454,641,311	140,431,866	73½
Colorado	Nov. 1, 1876.	None.	50,000	74,000	25,584,669	18,545,536	15
Connect't	Dec. 1, 1876.	5,014,500	.....	705,094	238,027,082	106,379,945	15
Delaware	.....	1,000,000 (?)	.....	.....	.....	.....	50
Florida	Jan. 1, 1877.	1,259,600	43,392	225,000	19,713,462	10,197,991	70
Georgia	Jan., 1877.	11,135,500	.....	.....	146,041,809	99,811,941	50
Illinois	Jan. 1, 1876.	796,330	None.	2,640,025	931,199,308	197,291,421	96
Indiana	Nov. 1, 1877.	1,097,736	.....	1,326,484	638,246,860	222,362,781	13
Iowa	Oct. 30, 1877.	545,435	$41,000	(?) 750,000	324,696,364	79,971,680	20
Kansas	Nov. 1, 1877.	1,235,975	.....	714,549	94,586,008	29,246,313	55
Kentucky	Oct. 10, 1876.	2,477,000	183,394	1,586,138	211,508,998	396,594,486	40
Louisiana	May 31, 1876.	9,318,343	2,548,812	2,473,629	189,220,457	35,465,377	14½
Maine	Jan. 1, 1877.	5,920,400	.....	675,173	224,579,569 (Real & Personal.) 547,044,271 (Real & Personal.)		80
Maryland	1877.	10,296,528*	.....	.....	.....	.....	17½
Massac'ts	Jan. 1, 1877.	33,550,464†	17,072	1,800,000	1,191,499,228	508,965,487	10

Data coverage

- A comprehensive dataset spanning 1863–2021 for 48 US states.

Categories	# of Variables	Examples
Labor Market	3	non-farm employment, manufacturing payroll
Real Activity	34	agricultural receipts, manufacturing value added, mining value, exports value
Housing	2	house sales price index, rent price index
Transportation	4	railroad mileage, motor vehicle registration
Government	9	federal govt revenue, state general revenue, state general expenditure, state debt
Wealth	4	personal income, value of farm land, # of bankruptcies
Others	4	# of patents, news-based sentiment index, newspaper circulation
Total	**60**	

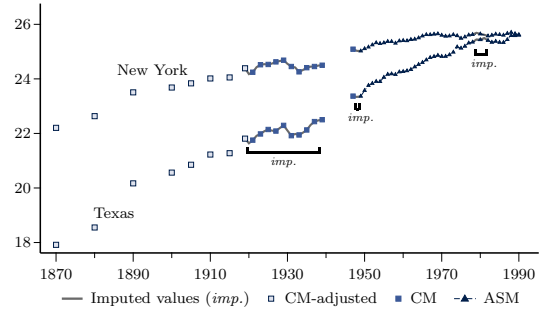
AGR MNM GOV BIZ OTH

Data challenges

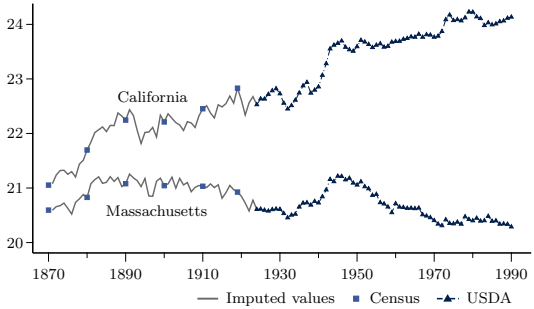
- What is not covered?
  - Data not always reported, e.g., state-level retail sales (4 observations only pre-1963)
- Poor document quality in earlier years
  - AWS OCR for digitization
  - Check accuracy by hand
- Frequent changes in variable definitions and data sources
  - Standard methods to harmonize across sources and definitions (e.g., chain-linking)
- Aggregate series not always available, but subcomponents often are
  - Chow-Lin method (e.g., agriculture, mining sector output) AGR
- All such adjustments are documented in an extensive data appendix

Examples of imputed and harmonized time series

(a) Value added of mfg. production (log-scale)



(b) Value of agric. products sold (log-scale)



Comparison with the existing literature

	Variable	Frequency	Coverage
A. State-Level			
This paper	Economic activity index	Annually	1871–2021
BEA	Personal income	Annually	1929–2024
BEA	GDP	Annually	1963–2024
Crone and Clayton-Matthews (2005)	Coincident index	Monthly	1978–2003
Baumeister, Leiva-León and Sims (2024)	Economic conditions index	Weekly	1987–2023
B. National-Level Historical Data			
Davis (2004)	Industrial production	Annually	1790–1915
Miron and Romer (1990)	Industrial production	Monthly	1884–1940
Federal Reserve	Industrial production	Monthly	1919–2023
Williamson (2025)	GDP	Annually	1790–2023
Balke and Gordon (1989)	GNP	Annually	1869–1929
BEA	GDP	Annually	1929–2023

Constructing an Index of State-Level Economic Activity

Dynamic factor model

- Let $Y_{j,t}$ denote the value of an indicator j in year t
  - If it is reported in years t and $t-1$, then we observe the year-on-year growth rate $y_{j,t} = \ln Y_{j,t} - \ln Y_{j,t-1}$
  - If it is reported in t and $t - \mathcal{T}_{j,t}$ for $\mathcal{T}_{j,t} > 1$, then $y_{j,t}$ is not observed
  - Instead, we observe

$$\tilde{y}_{j,t} = \frac{1}{\mathcal{T}_{j,t}} \Big(\ln Y_{j,t} - \ln Y_{j,(t-\mathcal{T}_{j,t})} \Big) = \frac{1}{\mathcal{T}_{j,t}} \Big(\underbrace{y_{j,t} + y_{j,t-1} + \cdots + y_{j,t-(\mathcal{T}_{j,t}+1)}}_{\text{these objects are not observed}} \Big) \tag{1}$$

- We postulate that $y_{j,t} = \lambda_j f_t + u_{j,t}$, where

$$f_t = \phi_1 f_{t-1} + \phi_2 f_{t-2} + \cdots + \phi_{l_f} f_{t-l_f} + \epsilon_t, \qquad \epsilon_t \sim \text{IID}\, \mathcal{N}(0, \sigma_f^2),$$
$$u_{j,t} = \psi_{j,1} u_{j,t-1} + \psi_{j,2} u_{j,t-2} + \cdots + \psi_{j,l_u} u_{j,t-l_u} + \varepsilon_{j,t} \qquad \varepsilon_{j,t} \sim \text{IID}\, \mathcal{N}(0, \sigma_j^2).$$

Dynamic factor model

- We can express $\tilde{y}_{j,t}$ in terms of f_t and $u_{j,t}$, such that

$$\begin{aligned}\tilde{y}_{j,t} &= \frac{1}{\mathcal{T}_{j,t}}\Big(y_{j,t} + y_{j,t-1} + \cdots + y_{j,t-(\mathcal{T}_{j,t}+1)}\Big) \\ &= \frac{1}{\mathcal{T}_{j,t}}\lambda_j\Big(f_t + f_{t-1} + \cdots + f_{t-(\mathcal{T}_{j,t}+1)}\Big) + \frac{1}{\mathcal{T}_{j,t}}\Big(u_{j,t} + u_{j,t-1} + \cdots + u_{j,t-(\mathcal{T}_{j,t}+1)}\Big)\end{aligned}$$

- We stack the growth rates of all indicators into a vector, $\mathbf{y}_t$, where

$$\underbrace{\mathbf{y}_t}_{m\times 1} = \underbrace{\mathbf{H}_t}_{m\times n} \cdot \underbrace{\boldsymbol{\alpha}_t}_{n\times 1} \tag{2}$$

  and α_t contains lag polynomials of common and unique factors (n is fixed for all t)

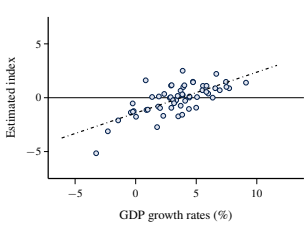
- For $t = t_0, \ldots, t_{end}$, $\boldsymbol{\alpha}_t$ evolves as follows

$$\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}) \tag{3}$$

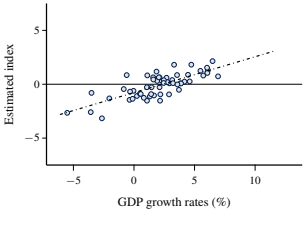
Estimation steps

- Let $\mathbf{Y} = [\mathbf{y}_{t_0}, \ldots, \mathbf{y}_{t_{end}}]$ and Θ be a vector containing all parameters
- Estimate the state space system based on the Gibbs sampler
  1. Draw $\boldsymbol{\alpha}_t$, conditional on $\mathbf{Y}$ and Θ
     - Derive the Kalman filter and smoother for the state space model
     - For each t, estimate $E(\boldsymbol{\alpha}_t|\mathbf{Y};\Theta)$
  2. Draw Θ, conditional on $\boldsymbol{\alpha}_t$ and $\mathbf{Y}$
     - Assume each set of parameters in Θ has normal or normal-inverse-Gamma priors
- Perform these two steps sequentially for N times
- Approximate the index of economic activity by $\tilde{f} = (\boldsymbol{\lambda}^\top \boldsymbol{\lambda})^{-1} \boldsymbol{\lambda}^\top \mathbf{y}^{\mathrm{P}}$
  - $\boldsymbol{\lambda}$ is the median estimates of the factor loadings
  - $\mathbf{y}^{\mathrm{P}}$ is the input data with missing observations replaced by the projected values of the Kalman filter
- Rescale $\tilde{f}$ so that the index has an average growth and variance as the real GSP post-1964

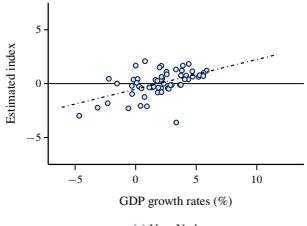
Validation: The Factor Estimates v.s. GDP from BEA



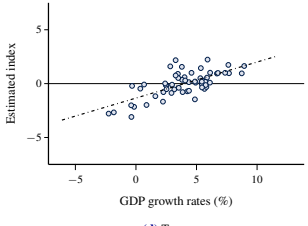
(a) California



(b) Illinois



(c) New York

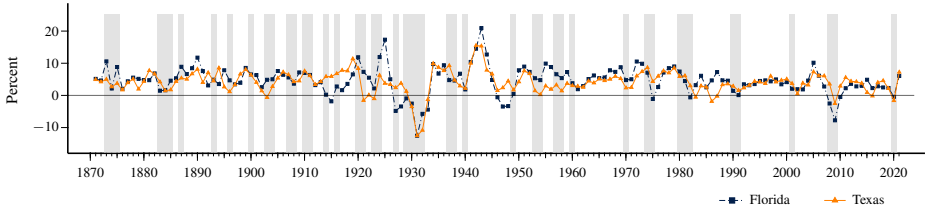


(d) Texas

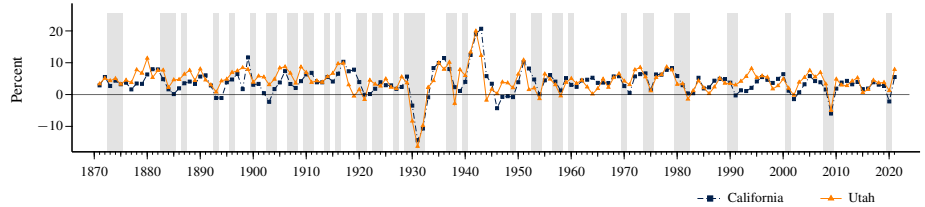
Index vs. Other Measures | Aggregated Index vs. US Measures | Assoc. with Vars. not used in Index | Alt. Inputs

State Economic Conditions Index (South and West)

North & Midwest

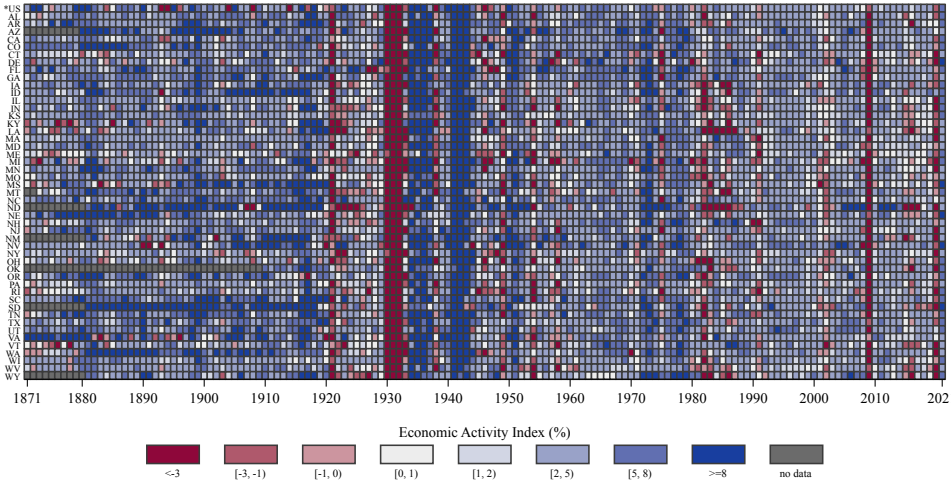


(a) South



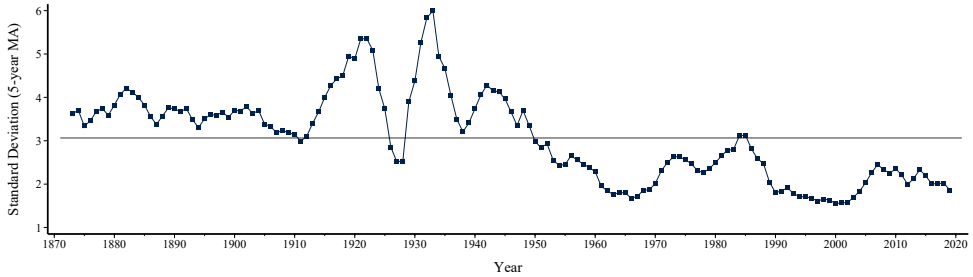
(b) West

State Economic Conditions Index: Heterogeneity across States



Dispersion of State-Level Economic Activities Over Time

Decomposition

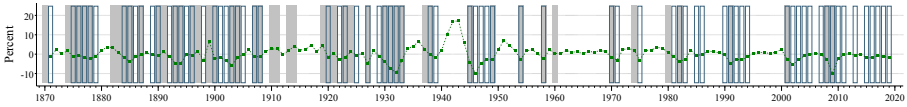


- State-level cycles have become more **synchronized** after World War II
- Dramatic decreases in standard deviation during Great Depression and 2007-08 Global Financial Crisis

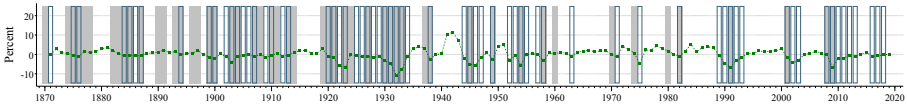
Identifying State-Level recessions (In Progress)

- We identify state-level recessions using by applying the workhorse Bry and Boschan (1971a) turning point method to our State-Level Economic Index
- State-level recessions tend to coincide with national recessions, but there are also exceptions:
  - local and U.S.-wide recessions are clearly correlated but distinct events

Identifying State-Level recessions (In Progress) Coincidence Rate



(a) California



(b) Massachusetts

Concluding remarks

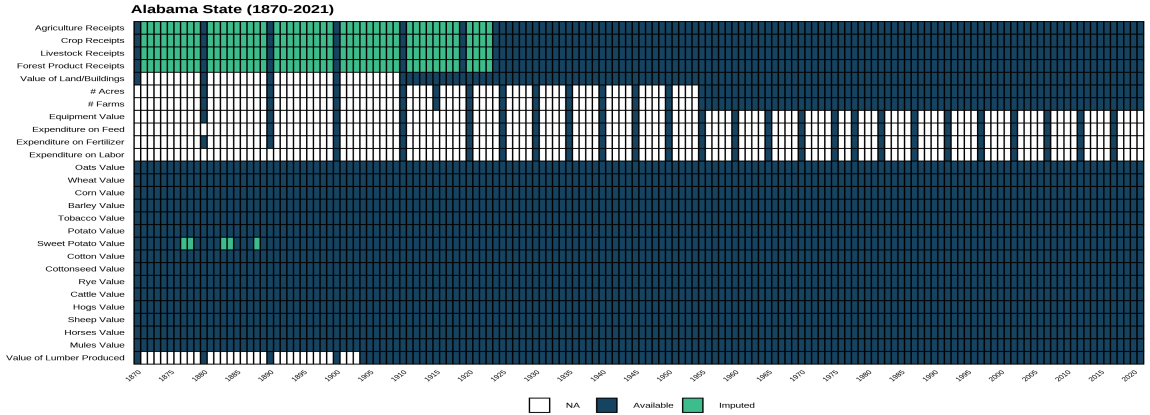
- We construct a new state-level panel dataset of economic indicators that ranges 1863–2021
- We implement a mixed-frequency dynamic factor model to construct an index of state-level economic conditions that covers over 150 years
  - Handles data inputs with mixed-reporting frequencies and arbitrarily missing patterns
  - Our estimated index tracks state-level GDP growth well
- We document substantial heterogeneity in state-level business cycles across space and over time
- In our view, these data will be useful for many other applications: Economic history, Growth, Empirical macro, Macro-finance
  - A different paper: Estimating the economic impact of financial crises, combining the state-level economic data with banking data

Thank you!

Additional Slides

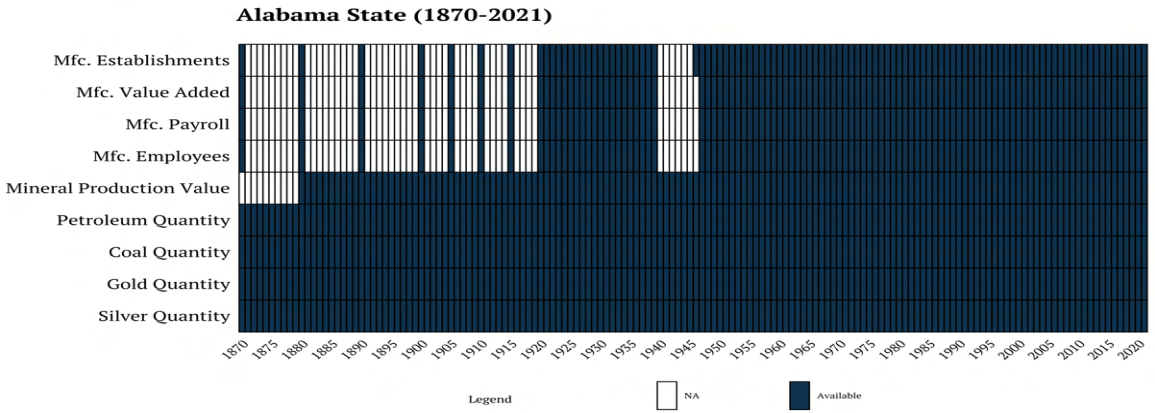
Agriculture

BACK



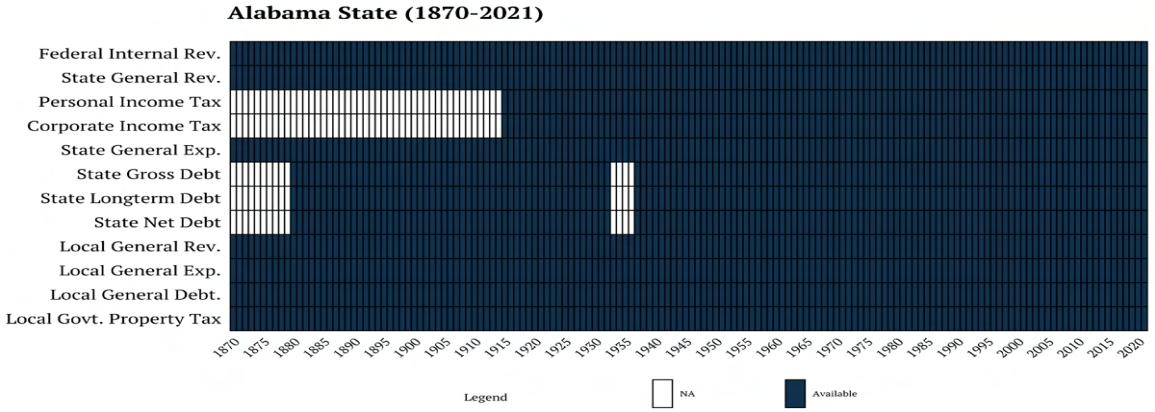
Manufacturing & Mining

BACK



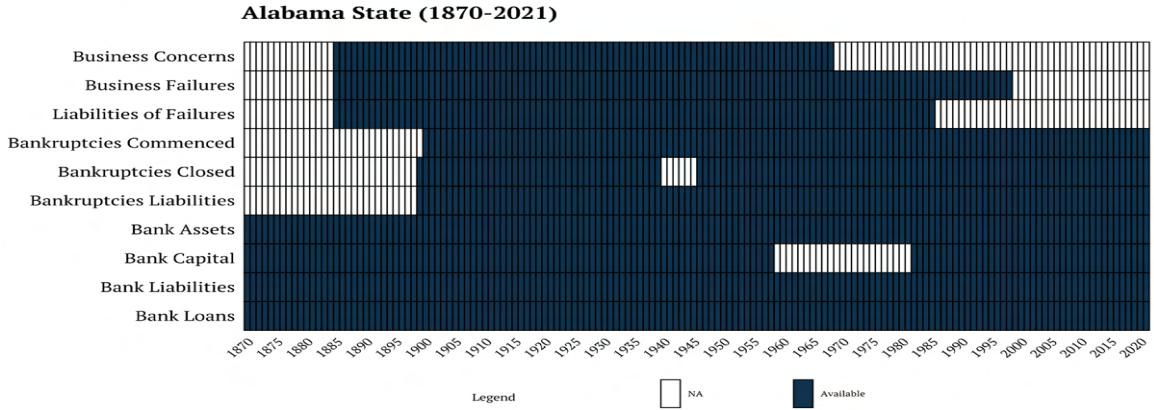
Government Finances

BACK

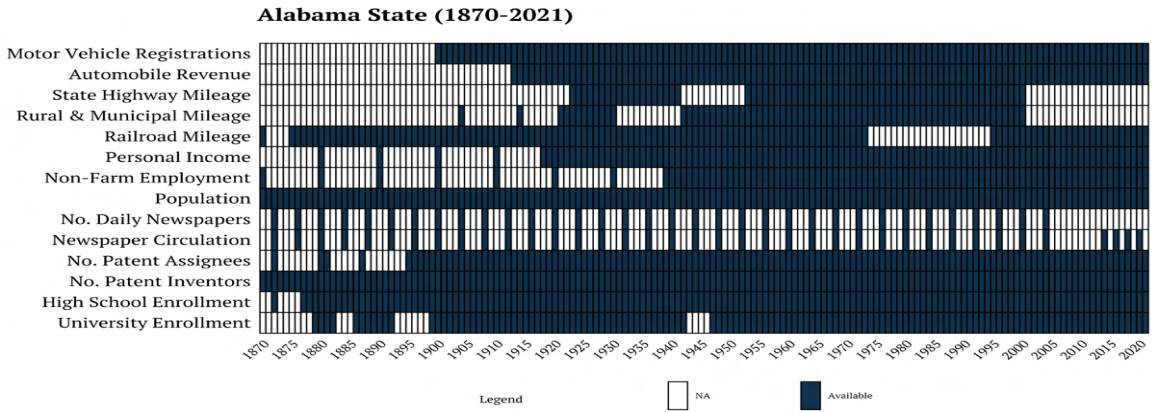


Business Statistics

BACK

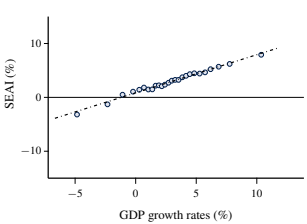


Labor market, Transport & Others BACK



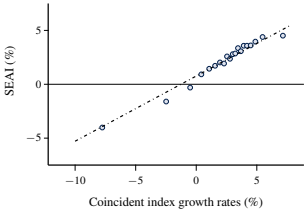
Validation: SEAI and Other Measures of Economic Conditions

Back

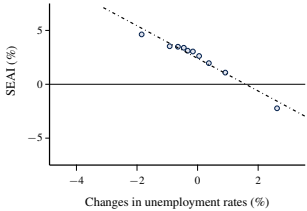


(a) GDP

(b) Personal income



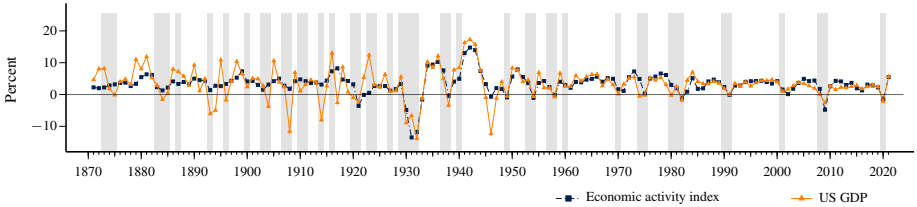
(c) Coincident index



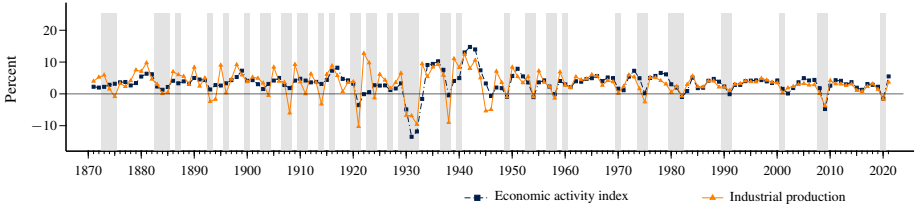
(d) Unemployment rates

Validation: Aggregates vs. U.S. Measures

Back



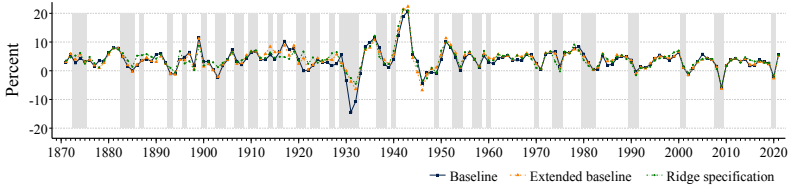
(a) Aggregated economic activity index and US GDP



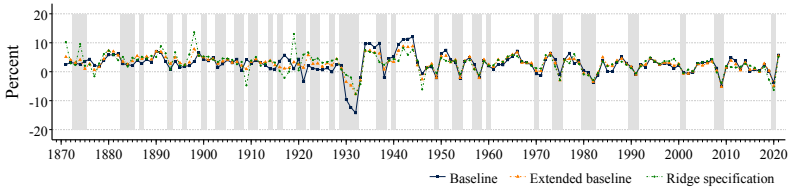
(b) Aggregated economic activity index and US industrial production

Validation: Alternative Estimates

Back



(a) California



(b) Illinois

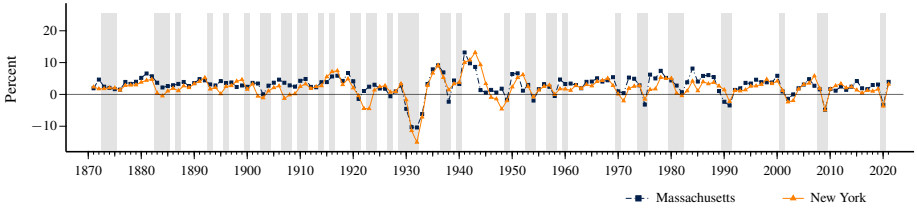
Validation: Variables not used in Index

Back

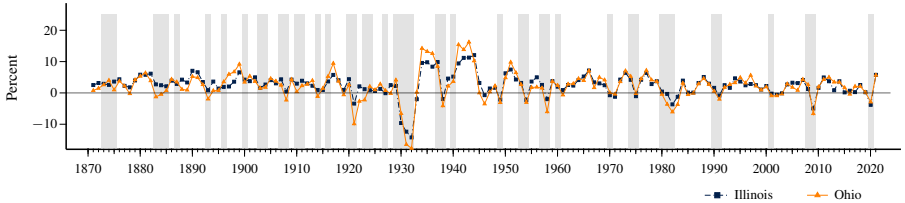
Indicator	$\hat{\beta}$	t-stat	Within-R^2
Manufacturing payroll	8.31^{***}	11.14	0.20
Number of manufacturing employees	9.97^{***}	11.64	0.28
Number of manufacturing establishments	4.67^{***}	9.90	0.02
Number of patents	1.07^{***}	3.80	0.00
Number of bankruptcies commenced	-3.33^{***}	-7.85	0.03
Number of bankruptcies terminated	-1.28^{***}	-5.25	0.00
Number of business failures	-4.98^{***}	-10.61	0.06
Total circulation of newspapers	3.55^{***}	4.70	0.03
Change in lagged sentiments	1.81^{***}	6.79	0.01

State Economic Conditions Index (Northeast and Midwest)

Back



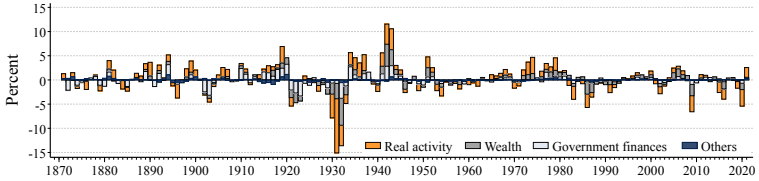
(a) Northeast



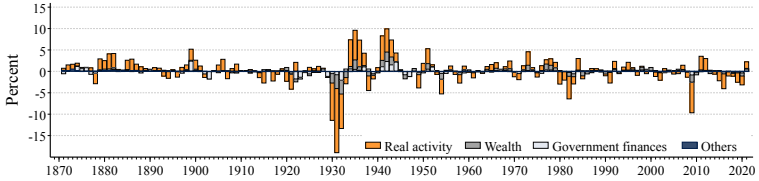
(b) Midwest

Decomposition of Estimated Index

Back

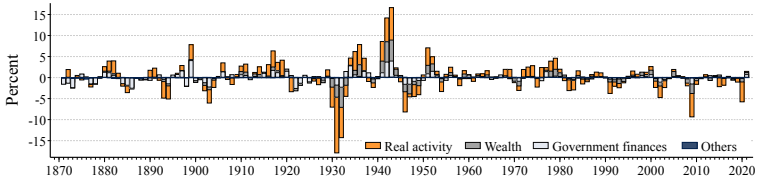


(a) Texas

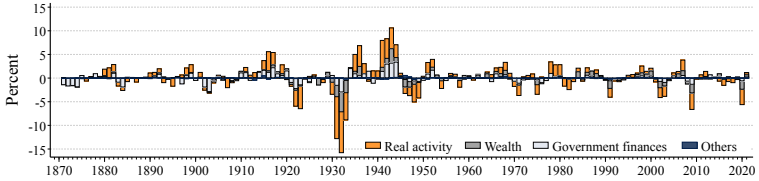


(b) Wisconsin

Decomposition of Estimated Index



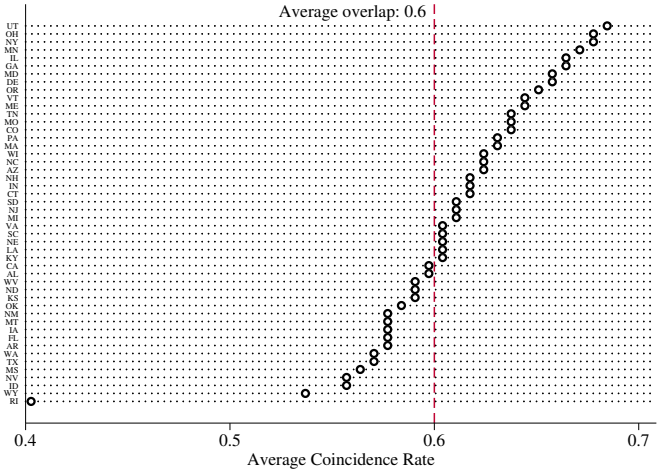
(a) California



(b) New York

Coincidence Rate by States

Back



- We calculate the coincidence rate as the fraction of times where a state and the U.S. as a whole are both