

The Role of Startups for Local Labor Markets: Online Appendix

Gerald Carlino and Thorsten Drautzburg*

May 20, 2020 May 20, 2020

*Federal Reserve Bank of Philadelphia, 10 Independence Mall, Philadelphia, PA, 19106. Phone: (215) 574-6000. Email: jerry.carlino@phil.frb.org and tdrautzburg@gmail.com. An earlier version of this paper circulated as “All jobs are created equal? Examining the importance of startups for local labor demand”. This paper benefited from comments by Jonás Arias, Marco Del Negro (editor), Karel Mertens (discussant), Jonathan Wright, and audiences at the Federal Reserve Bank of Philadelphia, the 2016 AEA meetings, and the ERSA meetings. Catherine O’Donnell, Adam Scavette, Samuel Wascher, and Nick Zarra provided excellent research assistance. Any mistakes are our own. Thorsten Drautzburg gratefully acknowledges the hospitality of the Federal Reserve Bank of Chicago and the Becker-Friedman-Institute during the final stage of this project. The views expressed herein are our own views only. They do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve System, or its Board of Governors.

Appendix

A Data

Startup data. We use the administrative data from the U.S. Census Bureau’s Business Dynamics Statistics (BDS), described in [Haltiwanger et al. \(2013\)](#), for sector-level and metropolitan area-level employment by firm age, and firm exit and entry rates.²⁷

Metropolitan area definitions. For consistency with the BDS, we aggregate counties to MSAs using the 2009 MSA definitions. In rare cases, the definitions of counties themselves have changed over time. We found only one MSA that was affected by this change.²⁸ Table [B.1](#) lists all MSAs in our sample.

Metropolitan employment. We use the administrative data from the Census’ County Business Patterns for average wage rates and industry-level employment at the MSA level. Average wage rates are simply first quarter payroll per employee, both summed across all counties within an MSA. To compute industry-level employment, we need to impute some county-industry-level employment data.

We build on [Autor et al. \(2013\)](#) for our imputation. Their code uses the county of establishments within industry-size brackets as well as employment totals at higher levels to impute county-level employment at the four digit SIC and six digit NAICS level. Intuitively, the algorithm computes a year-industry specific mapping of the binned size distribution of establishments to total year-industry employment. The algorithm runs repeatedly until estimates on the pooled disclosed and imputed data converge. We then aggregate the county-industry-level employment to the metropolitan level. Following [Autor et al. \(2013\)](#), we begin with OLS imputation, imposing lower and upper bounds for average employment by size bracket after the estimation. This procedure always converges for the decadal data analyzed in [Autor et al. \(2013\)](#), but not in all years. When the OLS analysis with ex-post bounds does not converge, we switch to non-linear least squares that imposes the bounds during the estimation. After imputing employment according to the prevailing classification scheme in each year, we use cross-walks from the 1977 SIC classification to the 1987 SIC classification and from future NAICS classifications to the 1997 six-digit NAICS classification, which we then, in turn, transform to the 1987 four-digit SIC classification and aggregate up to the three-digit level.²⁹ We also use these data to compute sectoral weights to predict startup activity.

Migration data. Migration data for the United States are obtained from the Internal Revenue Service’s (IRS) Statistics on Income Division. The migration data are based on year-to-year address changes reported on individual income tax returns filed with the IRS. We use county-to-county flows of

²⁷The sectors are: Agricultural services, mining, construction, manufacturing, utilities, wholesale, retail, FIRE (finance, insurance, and real estate), and services.

²⁸The newly created Broomfield county was split out of the Boulder, CO, MSA and as a new county became part of the neighboring Denver, CO, MSA. We therefore combine the data on the Denver and Boulder MSAs. In 1997, Dade county, FL, was renamed to Miami-Dade county. This change does not affect our analysis.

²⁹For the NAICS to SIC crosswalk, we use the crosswalk from [Autor et al. \(2013\)](#). We could not find a comprehensive crosswalk for the minor within-SIC and within-NAICS changes. To that end, we first use correspondence tables to identify the mapping between sectors. For some industries, this identifies the mapping uniquely, i.e., 100% of one or more industries map into a single industry. If one industry maps into more than one industry, we compute the weights in the crosswalk by regression: We regress the share in the originating industry in the last year of the old classification on the shares of the receiving industries in the new classification at the county-industry level. In our baseline specification, we use OLS and set negative coefficients to zero before normalizing weights to add up to unity. A non-linear LS procedure that respects these constraints yields similar results, but can become unwieldy in the rare cases when a large number of industries are the receiving industries, e.g., in the case for some wholesale sectors.

exemptions for the period 1990–2013 and convert the county flows to MSA flows using the 2009 MSA definitions. For the period from 1984 to 1989, we use archived IRS data from the National Archives.³⁰ As most households file tax returns by mid-April of each year, the migration data lines up with our snapshot of employment and startup data.

Population and density data. Population data were obtained at the county level from the Census Bureau. Counties are aggregated to MSAs using the 2009 MSA definitions. We calculate population density as population per square km per MSA, using data from the Census’ American Fact Finder.³¹ Population is measured for the middle of the calendar year.

MSA proximity matrix. We construct the proximity matrix based on the Census Bureau’s MAF/TIGER shape files. The proximity is the (inverse) squared Euclidean distance between the centroids of any pair of MSAs.³²

House prices. Metropolitan area housing prices are from the CoreLogic Solutions monthly repeat-sales Housing Price Index. We use March values of the index at the MSA level. However, MSA data were not available for eleven MSAs (Boston, Chicago, Dallas, Detroit, Los Angeles, Miami, NYC, Philadelphia, San Francisco, Seattle, Washington DC). For these MSAs, we use data at the NECTA metropolitan division level.³³

Main variable definitions.

- Net migration rate: We define the net migration rate as the difference between inflows and outflows of IRS exemptions, divided by the population level in the prior period. The number of exemptions on tax returns is, typically, the number of household members.

$$\text{net migration rate}_{m,t} = \frac{\text{No. exemptions (inflow)}_{m,t} - \text{No. exemptions (outflow)}_{m,t}}{\text{Population}_{t-1}} \quad (\text{A.1})$$

- Job creation rate: We define the job creation rate as the change in job creation by firms aged 0, divided by the average of overall private employment in the current and prior year.

$$\Delta \text{job creation rate}_{m,t} = \frac{\Delta \text{job creation by firms aged 0 in MSA}_{m,t}}{\frac{1}{2}(\text{MSA employment}_{m,t-1} + \text{MSA employment}_{m,t})} \quad (\text{A.2})$$

The numerator follows [Haltiwanger et al. \(2013\)](#).

³⁰See <https://www.irs.gov/uac/soi-tax-stats-migration-data> for data from 1990 onward and <https://catalog.archives.gov/id/646447> for the archived data.

³¹The population data is from <https://www.census.gov/programs-surveys/popest/data/data-sets.All.html> and the area data from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_09_5YR_G001&prodType=table.

³²<https://www.census.gov/geo/maps-data/data/tiger-line.html>.

³³We use CoreLogic Solutions’ Single Family Combined Index (HPI 4.0 Data) that excludes distress sales. There are 11 MSAs deemed large enough to be subdivided into their component metropolitan divisions. For example, the Dallas-Fort Worth-Arlington, Texas MSA is composed of the Dallas-Plano-Irving, Texas Metropolitan Division and the Fort Worth-Arlington, Texas Metropolitan Division. We aggregate both metropolitan divisions in this case, and proceed similarly in the other ten cases. We use house prices for the first quarter of each calendar year to line up approximately with the BDS and CBP data. In rare cases, the first quarter is missing so we use the last quarter of the preceding year.

- Δ Firm entry rate: We define the firm entry rate as the change in the number of firms aged 0, divided by the average of the number of firms of any age in the current and prior year.

$$\Delta \text{firm entry rate}_{m,t} = \frac{\Delta \text{Firms aged 0 in MSA}_{m,t}}{\frac{1}{2}(\text{All firms}_{m,t-1} + \text{All firms}_{m,t})} \quad (\text{A.3})$$

- Δ Firm exit rate: We define the firm exit rate as the change in the number of firms aged 1 that exit, divided by the average of the number of firms of any age in the current and prior year.

$$\Delta \text{firm exit rate}_{m,t} = \frac{\text{Firm deaths of firms aged 1 in MSA}_{m,t}}{\frac{1}{2}(\text{All firms}_{m,t-1} + \text{All firms}_{m,t})} \quad (\text{A.4})$$

- Δ Overall firm exit rate: We define the overall firm exit rate as the change in the number of firms of any age that exit, divided by the average of the number of firms of any age in the current and prior year.

$$\text{overall firm exit rate}_{m,t} = \frac{\Delta \text{Firm deaths of any firm in MSA}_{m,t}}{\frac{1}{2}(\text{All firms}_{m,t-1} + \text{All firms}_{m,t})} \quad (\text{A.5})$$

- Startup average size: We compute the average size of a startup as the log of the ratio of startup employment in an MSA divided by the number of startups in an MSA.
- Employment-to-population ratio: We use overall employment from the County Business Patterns to compute the log growth rate. This measure agrees closely with BDS employment; see Figure B.1. It enters the analysis in logs.
- Population growth: We compute the log growth rate. The log growth rate has the advantage of being additive to compute level changes, from which we can back out the change in the employment level.
- Growth of average wages: We compute the log growth rate of the average wage rate in the County Business Patterns.
- House price growth: We compute the log growth rate of first quarter house prices.
- TFP growth: See Appendix E.

Definition of instruments.

- Overall labor demand shock proxy:

$$Z_{m,t}^{\text{overall}} = \sum_i \omega_{m,i,t-5}^{\text{SIC3}} \Delta(\log(\text{emp}_{i,t} - \text{emp}_{m,i,t})) \quad (\text{A.6})$$

- Startup productivity shock proxy:

$$Z_{m,t}^{\text{startup}} = \sum_i \omega_{m,i,t-5}^{\text{sector}} \Delta \text{job creation rate}_{i,t} \quad (\text{A.7})$$

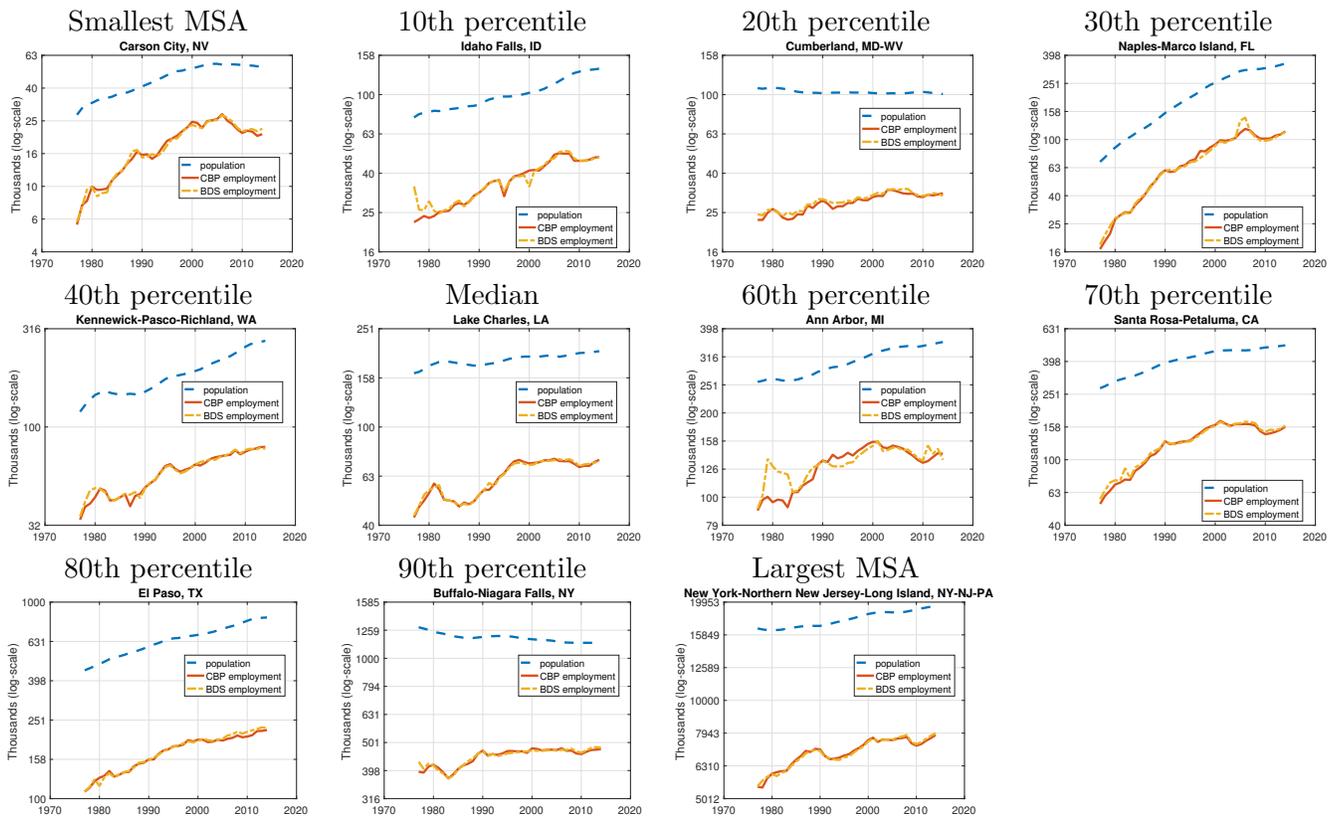
- Barriers to entry shock proxy:

$$Z_{m,t}^{\text{barriers}} = \sum_i \omega_{m,i,t-5}^{\text{sector}} \Delta \text{firm entry rate}_{i,t} \quad (\text{A.8})$$

- Overall TFP-based labor demand shock proxy, where the SIC classification follows Table [E.1](#):

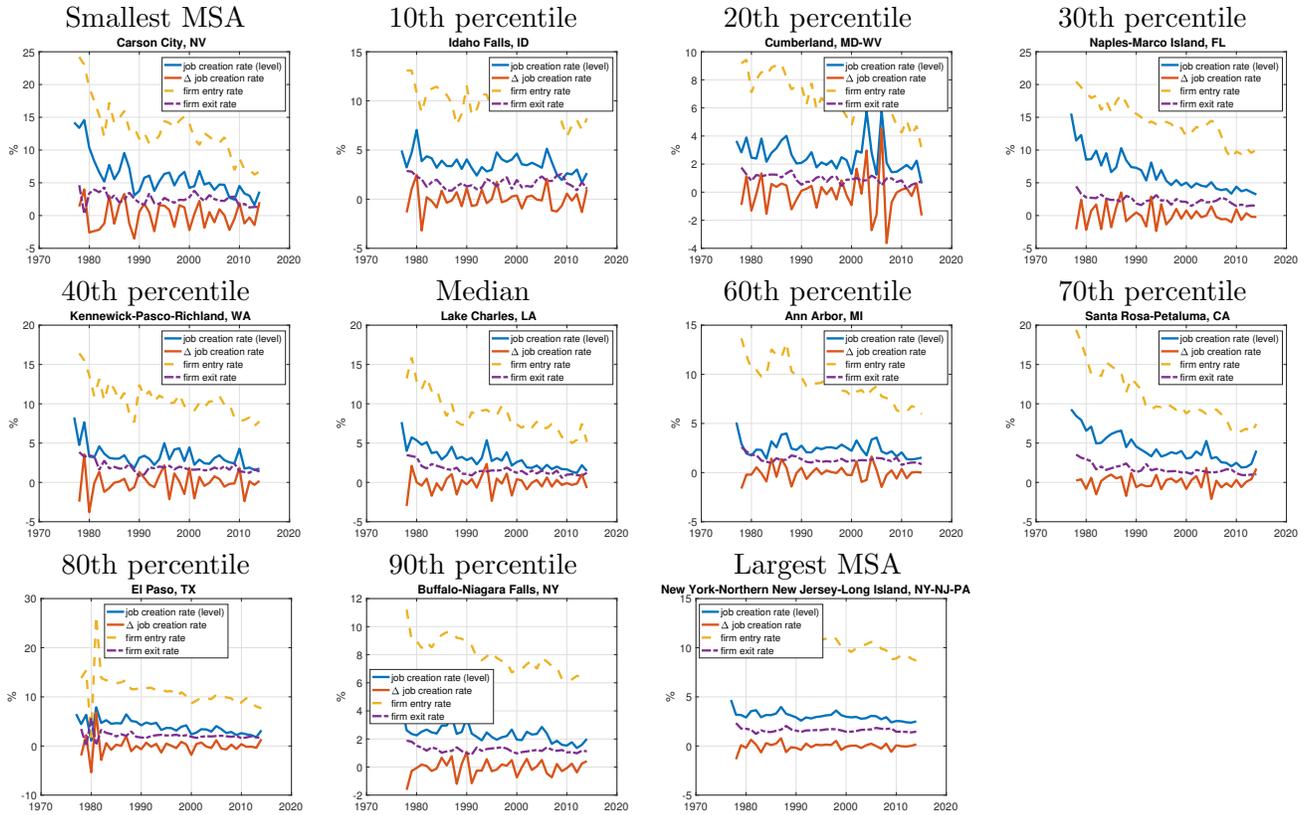
$$Z_{m,t}^{\text{overall}} = \sum_i \omega_{m,i,t-5}^{\text{SIC}} \Delta \log(TFP_{i,t}) \quad (\text{A.9})$$

B Data and descriptive statistics



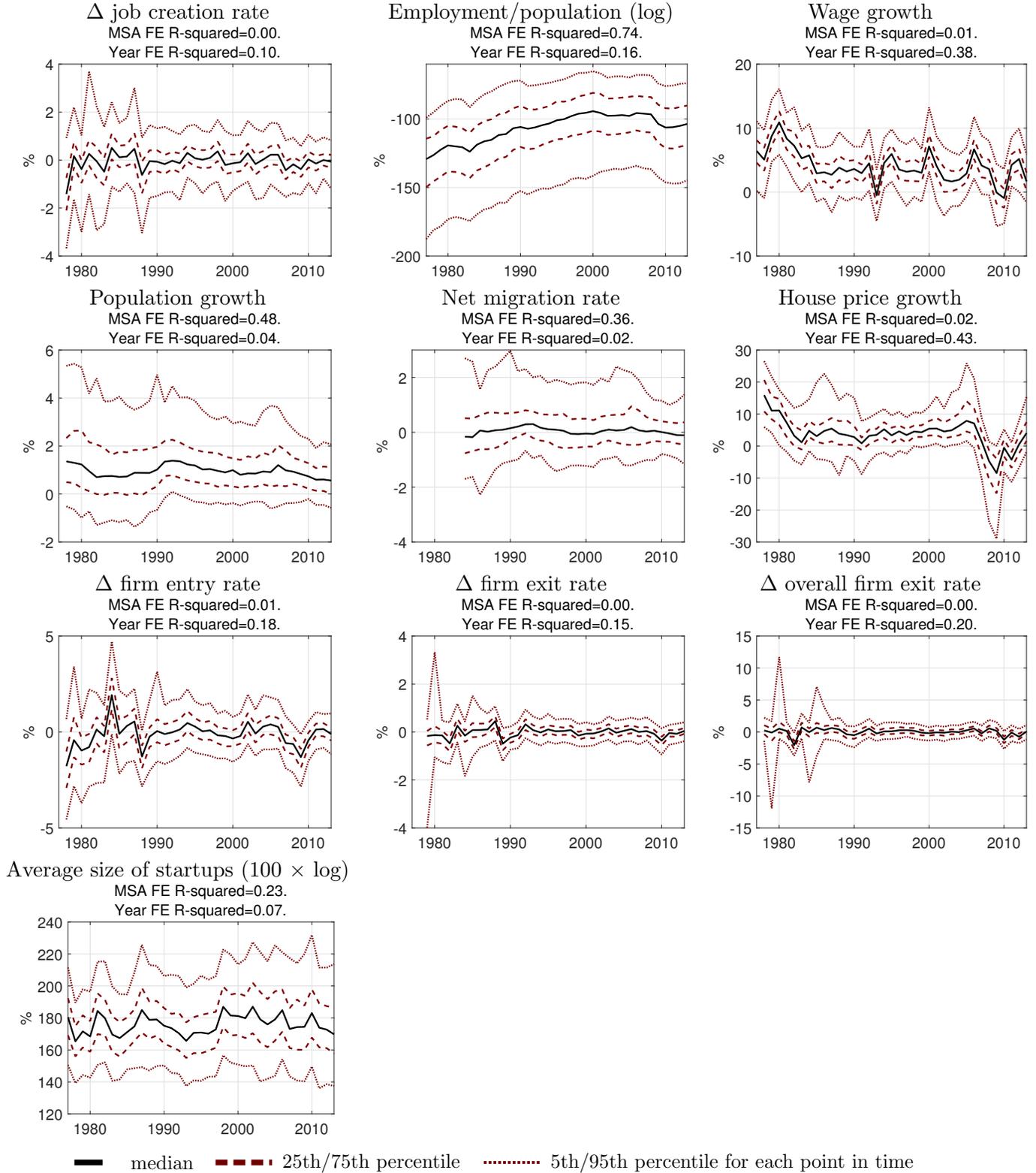
County Business Pattern (CBP) employment measures and Business Dynamics Statistics employment (BDS) track each other closely. Population levels change smoothly. We show the MSAs with the smallest and largest population in 1986, the start of the migration series, and MSAs next to the deciles of the population size distribution.

Figure B.1: Population, CBP employment, and BDS employment for MSAs of various sizes



Firm entry rates show a trend decline, whereas job creation rates are almost stable in many MSAs such as New York, NY, or Ann Arbor, MI. Focusing on the change in job creation by startups, converted to a rate relative to overall employment, we see stationary series. Firm exit rates also appear stationary. We show the MSAs with the smallest and largest population in 1986, the start of the migration series, and MSAs next to the deciles of the population size distribution.

Figure B.2: Startup activity for MSAs of various sizes



The variables in our VAR and its periphery have regional variation that we model and use for identification. We show the median across MSAs along with the inner 50% and 90% for each point in time. House price data are from CoreLogic Solutions.

Figure B.3: Cross-sectional distribution of variables in VAR over time

MSA, state(s)	Initial density	Entry rate (% of firms)		Avg. startup size	
	Pop. per sq mi	Initial	2013	Initial	2013
Abilene, TX	49	15.0	6.1	5.3	6.9
Akron, OH	736	11.8	5.6	5.8	5.4
Albany, GA	73	12.5	5.1	5.0	6.6
Albany-Schenectady-Troy, NY	276	10.5	7.0	4.4	4.7
Albuquerque, NM	52	16.0	6.9	5.7	6.1
Alexandria, LA	75	11.7	4.4	5.9	12.8
Allentown-Bethlehem-Easton, PA-NJ	432	10.6	5.9	5.0	6.2
Altoona, PA	261	8.8	5.0	5.9	7.8
Amarillo, TX	47	14.6	6.5	5.4	4.6
Ames, IA	122	13.1	4.9	6.6	15.2
Anchorage, AK	7	21.4	7.1	5.2	5.3
Ann Arbor, MI	365	13.7	6.6	5.4	5.0
Anniston-Oxford, AL	186	14.6	4.4	5.4	4.6
Appleton, WI	161	12.2	4.9	5.0	5.4
Asheville, NC	133	14.8	7.5	4.6	4.9
Athens-Clarke County, GA	105	14.3	7.1	5.0	5.9
Atlanta-Sandy Springs-Marietta, GA	262	16.6	9.2	5.4	5.1
Atlantic City-Hammonton, NJ	341	12.9	6.2	4.0	6.8
Auburn-Opelika, AL	118	13.2	8.2	5.9	8.5
Augusta-Richmond County, GA-SC	109	13.6	5.6	5.6	10.7
Austin-Round Rock-San Marcos, TX	126	15.8	10.7	5.7	5.3
Bakersfield-Delano, CA	46	17.2	7.3	5.2	5.1
Baltimore-Towson, MD	846	11.9	6.9	4.6	5.6
Bangor, ME	40	13.3	5.3	3.2	5.0
Barnstable Town, MA	351	15.5	6.4	4.0	4.1
Baton Rouge, LA	136	13.3	6.3	6.0	6.2
Battle Creek, MI	202	10.1	4.6	5.5	5.4
Bay City, MI	275	10.3	4.2	5.1	5.5
Beaumont-Port Arthur, TX	173	12.1	5.9	6.1	5.6
Bellingham, WA	47	19.8	6.9	11.6	4.1
Bend, OR	16	27.4	8.3	5.2	3.6
Billings, MT	24	13.5	6.5	5.0	4.2
Binghamton, NY	219	10.8	5.4	5.4	5.3
Birmingham-Hoover, AL	172	14.0	6.2	5.8	9.7
Bismarck, ND	21	17.1	7.5	5.7	5.2
Blacksburg-Christiansburg-Radford, VA	117	14.6	5.3	7.0	7.7
Bloomington, IN	106	13.3	5.6	5.3	7.0
Boise City-Nampa, ID	22	16.7	8.6	5.3	5.2
Boston-Cambridge-Quincy, MA-NH	1132	11.3	6.9	5.6	5.2
Bowling Green, KY	92	12.8	7.9	8.7	5.1
Bremerton-Silverdale, WA	319	18.9	7.6	4.5	4.8
Bridgeport-Stamford-Norwalk, CT	1297	11.8	7.1	4.6	5.5
Brownsville-Harlingen, TX	219	13.8	7.0	6.0	5.9
Brunswick, GA	53	16.4	6.7	5.5	5.3
Buffalo-Niagara Falls, NY	824	11.2	6.4	5.2	6.1
Burlington, NC	232	11.3	5.2	3.9	5.1
Burlington-South Burlington, VT	118	16.4	5.7	7.0	4.5
Canton-Massillon, OH	414	12.3	5.1	5.1	4.7
Cape Coral-Fort Myers, FL	222	21.4	10.6	5.8	4.5
Cape Girardeau-Jackson, MO-IL	54	13.2	6.0	4.4	4.3
Carson City, NV	189	24.2	6.2	5.4	3.4
Casper, WY	12	15.6	5.6	5.9	4.8
Cedar Rapids, IA	105	11.9	5.2	5.5	4.5
Champaign-Urbana, IL	105	14.2	5.8	5.9	5.8
Charleston, WV	129	12.3	4.4	4.6	7.4
Charleston-North Charleston-Summerville, SC	157	15.5	8.1	6.3	5.4
Charlotte-Gastonia-Rock Hill, NC-SC	262	13.8	8.7	5.2	5.3
Charlottesville, VA	74	13.9	5.8	4.3	5.0
Chattanooga, TN-GA	197	13.1	6.5	5.7	5.5
Cheyenne, WY	25	16.6	8.8	5.3	4.8
Chicago-Joliet-Naperville, IL-IN-WI	1118	12.0	7.8	6.0	5.1
Chico, CA	79	17.0	7.0	4.0	4.1
Cincinnati-Middletown, OH-KY-IN	393	11.6	6.3	5.1	6.3
Clarksville, TN-KY	77	13.0	6.4	4.3	5.5
Cleveland, TN	99	13.0	4.5	4.1	4.9
Cleveland-Elyria-Mentor, OH	1108	10.8	5.5	6.1	7.6
Coeur d'Alene, ID	41	22.1	7.3	5.0	4.5
College Station-Bryan, TX	50	13.5	7.1	5.4	6.5
Colorado Springs, CO	115	18.0	8.0	4.9	4.1
Columbia, MO	91	14.3	7.8	5.5	3.9
Columbia, SC	129	14.9	6.6	4.7	5.7
Columbus, GA-AL	131	12.5	5.8	4.9	4.6
Columbus, IN	153	11.9	4.2	5.6	7.3
Columbus, OH	318	12.6	7.3	5.0	7.0
Corpus Christi, TX	182	12.8	6.2	5.7	9.7
Corvallis, OR	97	19.2	5.9	6.7	4.0
Crestview-Fort Walton Beach-Destin, FL	116	16.9	8.0	4.2	5.2
Cumberland, MD-WV	143	9.2	4.7	4.4	8.8
Dallas-Fort Worth-Arlington, TX	311	15.7	9.5	6.1	6.7
Dalton, GA	129	17.1	4.9	6.5	5.1
Danville, IL	109	10.8	3.4	4.5	5.1
Danville, VA	110	10.7	4.1	4.1	5.5
Davenport-Moline-Rock Island, IA-IL	176	12.1	5.3	5.8	5.3
Dayton, OH	484	12.0	5.2	5.2	5.9
Decatur, AL	92	15.7	5.8	5.3	7.2
Decatur, IL	226	11.3	4.9	5.0	6.8
Deltona-Daytona Beach-Ormond Beach, FL	208	17.0	8.1	5.0	4.1
Denver-Aurora-Broomfield, CO	182	16.0	9.3	5.8	4.9
Des Moines-West Des Moines, IA	135	13.6	6.3	5.9	5.3
Detroit-Warren-Livonia, MI	1119	12.8	7.5	6.1	6.7
Dothan, AL	65	14.5	5.3	6.4	7.2
Dover, DE	164	11.9	7.3	4.5	4.5
Dubuque, IA	155	11.5	4.7	5.6	3.7
Duluth, MN-WI	35	11.6	4.1	6.0	6.0
Durham-Chapel Hill, NC	160	13.3	7.0	5.2	5.0

Table B.1: List of MSAs included (continued on the next page...)

MSA, state(s)	Initial density	Entry rate (% of firms)		Avg. startup size	
	Pop. per sq mi	Initial	2013	Initial	2013
Eau Claire, WI	77	13.9	6.0	4.8	7.9
El Centro, CA	21	14.2	5.7	4.8	3.8
Elizabethtown, KY	111	15.4	7.8	3.8	4.0
Elkhart-Goshen, IN	288	11.3	5.4	6.2	6.9
Elmira, NY	244	10.5	4.7	4.6	4.2
El Paso, TX	444	13.8	8.0	6.0	5.2
Erie, PA	350	9.7	3.7	4.7	5.6
Eugene-Springfield, OR	55	20.0	6.3	5.3	6.5
Evansville, IN-KY	137	13.2	5.6	5.0	5.7
Fairbanks, AK	8	20.5	7.2	6.3	5.6
Fargo, ND-MN	47	13.2	6.8	4.8	4.4
Farmington, NM	13	15.9	5.3	6.4	13.8
Fayetteville, NC	252	14.0	6.0	4.6	6.4
Fayetteville-Springdale-Rogers, AR-MO	60	16.2	8.5	4.9	4.9
Flagstaff, AZ	4	14.4	6.5	4.8	6.5
Flint, MI	701	12.3	5.7	5.8	5.9
Florence, SC	122	12.9	5.2	5.1	5.5
Florence-Muscle Shoals, AL	103	13.6	4.9	4.3	7.6
Fond du Lac, WI	122	11.2	3.8	4.3	5.0
Fort Collins-Loveland, CO	51	20.4	8.7	4.8	4.8
Fort Smith, AR-OK	52	12.0	5.7	4.8	6.6
Fort Wayne, IN	249	11.6	5.4	4.9	7.0
Fresno, CA	82	16.2	6.8	5.6	5.2
Gadsden, AL	187	12.0	3.7	8.1	7.5
Gainesville, FL	116	17.4	7.8	5.7	6.1
Gainesville, GA	181	12.1	7.5	4.8	4.9
Glens Falls, NY	65	11.9	5.6	3.6	5.9
Goldsboro, NC	171	10.7	5.0	4.6	4.6
Grand Forks, ND-MN	30	11.3	5.0	4.9	9.0
Grand Junction, CO	21	17.5	5.9	5.8	3.6
Grand Rapids-Wyoming, MI	199	11.9	6.0	5.5	6.6
Great Falls, MT	32	13.6	5.5	5.5	5.6
Greeley, CO	28	13.9	9.2	4.7	4.2
Green Bay, WI	117	12.1	5.4	7.7	5.9
Greensboro-High Point, NC	239	13.1	6.2	5.5	4.9
Greenville, NC	105	14.4	5.9	4.7	5.4
Greenville-Mauldin-Easley, SC	201	13.3	6.9	5.2	6.2
Gulfport-Biloxi, MS	124	14.3	5.4	6.7	6.5
Hagerstown-Martinsburg, MD-WV	164	10.3	5.0	4.2	5.6
Hanford-Corcoran, CA	52	13.6	5.1	4.5	4.8
Harrisburg-Carlisle, PA	268	10.0	5.5	4.7	8.9
Harrisonburg, VA	87	11.5	6.0	4.4	5.7
Hartford-West Hartford-East Hartford, CT	687	11.5	5.8	4.8	6.7
Hattiesburg, MS	59	12.8	6.4	5.6	7.6
Hickory-Lenoir-Morganton, NC	157	13.1	4.8	5.5	5.6
Hinesville-Fort Stewart, GA	38	24.4	5.7	5.1	9.9
Hot Springs, AR	100	13.0	5.6	6.0	4.5
Houma-Bayou Cane-Thibodaux, LA	72	13.3	4.5	5.6	6.1
Houston-Sugar Land-Baytown, TX	318	17.0	9.6	6.3	7.1
Huntington-Ashland, WV-KY-OH	174	12.4	5.2	4.1	7.0
Huntsville, AL	174	14.6	6.6	4.7	5.7
Idaho Falls, ID	26	13.1	7.1	4.9	3.8
Indianapolis-Carmel, IN	308	13.1	7.6	5.6	5.8
Iowa City, IA	84	14.5	6.5	6.0	5.3
Ithaca, NY	182	13.8	5.0	4.4	4.7
Jackson, MI	214	9.5	4.2	5.0	8.7
Jackson, MS	105	14.9	6.8	5.2	4.9
Jackson, TN	100	11.9	5.3	5.7	7.3
Jacksonville, FL	220	14.8	9.5	6.0	4.8
Jacksonville, NC	153	15.5	6.5	4.5	6.6
Janesville, WI	192	11.8	4.3	5.3	6.4
Jefferson City, MO	47	13.4	5.5	4.9	4.5
Johnson City, TN	175	12.1	4.8	4.4	6.1
Johnstown, PA	276	9.0	3.9	5.6	7.0
Jonesboro, AR	61	14.2	6.9	4.8	6.0
Joplin, MO	98	12.2	6.7	4.9	3.7
Kalamazoo-Portage, MI	233	12.1	4.9	5.0	5.2
Kankakee-Bradley, IL	150	11.1	5.1	6.0	5.1
Kansas City, MO-KS	189	14.0	7.9	5.7	5.6
Kennewick-Pasco-Richland, WA	41	16.4	7.2	5.0	4.0
Killeen-Temple-Fort Hood, TX	83	13.2	6.8	6.1	6.1
Kingsport-Bristol-Bristol, TN-VA	133	13.6	5.2	5.2	5.2
Kingston, NY	140	11.1	6.0	3.6	4.6
Knoxville, TN	260	15.2	6.1	5.4	5.9
Kokomo, IN	188	10.5	4.3	5.4	6.2
La Crosse, WI-MN	106	12.1	5.0	6.2	7.3
Lafayette, LA	175	15.3	7.0	5.4	6.9
Lake Charles, LA	70	13.2	7.5	5.1	6.0
Lake Havasu City-Kingman, AZ	3	20.3	7.3	4.4	4.5
Lakeland-Winter Haven, FL	165	15.6	7.4	4.7	6.3
Lancaster, PA	372	11.0	5.7	4.9	4.8
Lansing-East Lansing, MI	238	12.7	5.9	5.3	5.1
Laredo, TX	27	12.5	9.0	5.9	4.5
Las Cruces, NM	23	15.6	5.8	6.2	5.7
Las Vegas-Paradise, NV	49	17.6	12.0	7.9	7.0
Lawrence, KS	138	13.6	5.1	6.0	4.5
Lawton, OK	108	12.4	4.7	5.3	8.5
Lebanon, PA	299	8.5	4.3	4.2	5.1
Lewiston, ID-WA	32	12.5	5.3	4.6	3.8
Lewiston-Auburn, ME	210	11.2	5.0	4.9	9.7
Lexington-Fayette, KY	210	14.0	7.2	5.6	4.7
Lima, OH	274	9.4	3.7	4.7	5.4
Lincoln, NE	142	13.4	6.5	5.0	4.9
Little Rock-North Little Rock-Conway, AR	115	13.9	7.1	5.2	6.0
Logan, UT-ID	33	14.5	6.9	5.5	3.6

Table B.1: List of MSAs included (continued on the next page...)

MSA, state(s)	Initial density	Entry rate (% of firms)		Avg. startup size	
	Pop. per sq mi	Initial	2013	Initial	2013
Longview, TX	87	14.3	6.0	6.0	5.3
Longview, WA	65	15.8	5.8	4.9	8.7
Los Angeles-Long Beach-Santa Ana, CA	1869	16.5	9.2	6.1	6.0
Louisville/Jefferson County, KY-IN	254	12.0	6.3	6.0	6.5
Lubbock, TX	118	14.0	6.5	5.1	6.1
Lynchburg, VA	88	11.9	5.7	4.9	3.6
Macon, GA	114	11.9	5.7	4.0	6.2
Madera-Chowchilla, CA	25	17.3	6.3	4.6	4.3
Madison, WI	137	13.1	6.1	5.5	5.3
Manchester-Nashua, NH	300	15.7	6.1	4.8	5.2
Manhattan, KS	58	12.7	5.2	4.3	7.9
Mankato-North Mankato, MN	65	10.2	4.1	5.7	6.0
Mansfield, OH	264	9.4	4.3	4.0	5.3
McAllen-Edinburg-Mission, TX	165	13.4	9.4	6.4	5.9
Medford, OR	44	20.1	6.6	5.4	4.5
Memphis, TN-MS-AR	211	13.0	6.4	5.8	6.4
Merced, CA	65	13.6	5.8	4.2	5.5
Miami-Fort Lauderdale-Pompano Beach, FL	579	19.6	11.4	6.0	4.8
Michigan City-La Porte, IN	181	11.0	3.7	4.9	3.8
Midland, TX	82	14.4	9.4	6.3	6.9
Milwaukee-Waukesha-West Allis, WI	958	11.5	6.0	5.1	6.6
Minneapolis-St. Paul-Bloomington, MN-WI	352	13.9	7.1	6.2	5.8
Missoula, MT	27	17.7	6.2	5.4	5.4
Mobile, AL	287	15.7	5.0	5.7	6.6
Modesto, CA	165	18.6	7.3	5.1	4.6
Monroe, LA	104	12.8	5.9	4.8	6.8
Monroe, MI	237	11.8	5.3	4.4	4.4
Montgomery, AL	101	15.5	5.7	6.6	7.0
Morgantown, WV	101	12.4	5.9	4.0	10.0
Morristown, TN	131	12.8	6.0	5.3	6.4
Mount Vernon-Anacortes, WA	33	19.1	5.4	8.7	3.8
Muncie, IN	332	10.3	4.2	6.6	6.4
Muskegon-Norton Shores, MI	315	11.0	4.6	4.9	5.1
Myrtle Beach-North Myrtle Beach-Conway, SC	83	18.6	8.1	5.9	4.8
Napa, CA	129	15.2	6.9	4.1	6.5
Naples-Marco Island, FL	35	20.4	9.5	5.3	5.0
Nashville-Davidson-Murfreesboro-Franklin, TN	151	15.4	8.4	5.3	6.7
New Haven-Milford, CT	1254	11.2	5.8	4.7	9.3
New Orleans-Metairie-Kenner, LA	417	12.4	6.7	6.4	7.1
New York-Northern New Jersey-Long Island, NY-NJ-PA	2480	11.5	8.9	4.9	5.0
Niles-Benton Harbor, MI	308	11.4	4.0	6.7	8.3
North Port-Bradenton-Sarasota, FL	236	18.4	9.3	5.6	5.7
Norwich-New London, CT	363	11.2	5.9	4.1	5.8
Ocala, FL	69	17.2	8.1	4.8	5.4
Ocean City, NJ	311	15.6	6.1	3.4	2.6
Odessa, TX	117	14.5	8.2	6.3	5.8
Ogden-Clearfield, UT	182	16.4	8.2	6.0	5.4
Oklahoma City, OK	149	14.4	7.7	5.8	6.4
Olympia, WA	148	19.4	7.0	5.4	3.9
Omaha-Council Bluffs, NE-IA	151	12.4	6.8	5.6	5.3
Orlando-Kissimmee-Sanford, FL	209	18.7	10.8	5.3	5.9
Oshkosh-Neenah, WI	301	11.8	4.7	7.1	5.1
Owensboro, KY	112	12.4	4.9	5.3	6.1
Oxnard-Thousand Oaks-Ventura, CA	260	18.3	7.9	5.1	5.1
Palm Bay-Melbourne-Titusville, FL	238	20.6	8.7	5.6	3.7
Panama City-Lynn Haven-Panama City Beach, FL	123	15.8	6.6	4.9	5.0
Parkersburg-Marietta-Vienna, WV-OH	121	12.0	4.3	4.2	6.8
Pensacola-Ferry Pass-Brent, FL	171	15.3	7.2	5.1	4.9
Peoria, IL	155	11.4	4.6	5.1	5.5
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	1144	11.5	7.0	5.4	6.9
Phoenix-Mesa-Glendale, AZ	97	17.8	9.5	6.0	6.2
Pine Bluff, AR	53	11.8	5.1	5.0	6.8
Pittsburgh, PA	508	10.8	5.3	4.7	6.0
Pittsfield, MA	156	11.8	4.2	4.3	4.7
Pocatello, ID	27	12.2	5.5	5.7	3.6
Portland-South Portland-Biddeford, ME	179	13.8	6.2	4.8	4.4
Portland-Vancouver-Hillsboro, OR-WA	186	17.1	8.2	5.2	4.4
Port St. Lucie, FL	114	20.0	9.6	5.9	4.8
Prescott, AZ	7	20.3	7.3	4.3	4.4
Providence-New Bedford-Fall River, RI-MA	896	10.8	6.0	6.1	5.1
Provo-Orem, UT	36	16.7	11.2	6.0	6.2
Pueblo, CO	52	14.2	6.5	4.5	4.9
Punta Gorda, FL	69	21.9	8.9	5.4	4.0
Racine, WI	516	11.4	5.3	8.5	4.7
Raleigh-Cary, NC	178	15.2	8.8	4.6	5.0
Rapid City, SD	15	17.8	7.0	5.7	4.8
Reading, PA	358	9.4	5.4	5.5	6.7
Redding, CA	27	21.9	5.3	4.1	5.6
Reno-Sparks, NV	26	17.4	7.4	7.1	5.7
Richmond, VA	144	13.2	7.2	5.0	5.3
Riverside-San Bernardino-Ontario, CA	49	19.4	9.3	5.5	5.8
Roanoke, VA	138	12.0	5.0	4.6	6.3
Rochester, MN	76	12.9	6.1	4.6	4.8
Rochester, NY	334	11.6	6.6	5.1	5.0
Rockford, IL	345	11.4	5.3	5.5	6.1
Rocky Mount, NC	116	12.1	5.1	4.3	6.5
Rome, GA	158	12.4	5.9	6.3	8.3
Sacramento-Arden-Arcade-Roseville, CA	196	18.9	8.3	5.5	5.1
Saginaw-Saginaw Township North, MI	284	11.2	4.5	5.0	7.0
St. Cloud, MN	74	13.3	5.1	5.5	4.6
St. George, UT	9	15.1	10.8	5.8	5.4
St. Joseph, MO-KS	73	12.3	7.1	4.8	6.3
St. Louis, MO-IL	291	12.1	8.7	5.6	4.5
Salem, OR	119	18.2	6.6	4.8	5.9
Salinas, CA	86	16.0	5.8	4.8	5.2

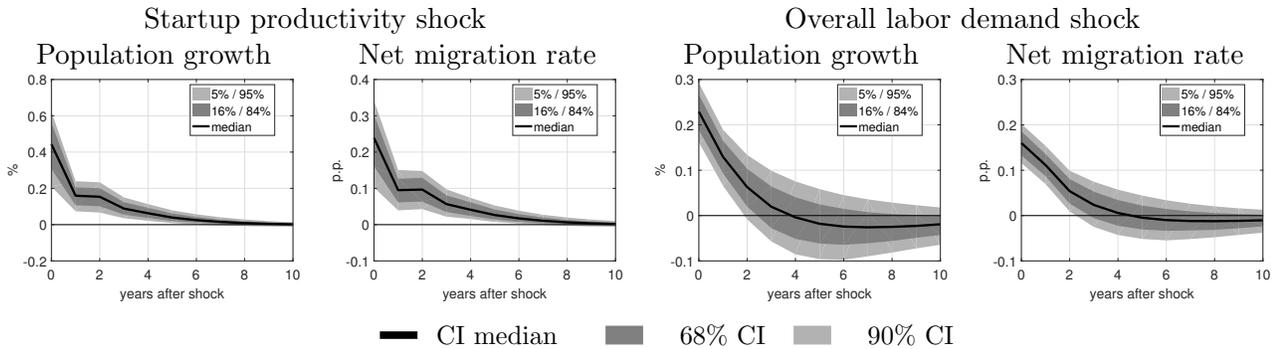
Table B.1: List of MSAs included (continued on the next page...)

MSA, state(s)	Initial density	Entry rate (% of firms)		Avg. startup size	
	Pop. per sq mi	Initial	2013	Initial	2013
Salisbury, MD	119	12.7	5.8	3.7	5.0
Salt Lake City, UT	62	15.3	8.9	6.2	5.5
San Angelo, TX	31	12.8	6.5	5.6	4.8
San Antonio-New Braunfels, TX	151	13.8	8.6	5.9	6.6
San Diego-Carlsbad-San Marcos, CA	408	20.9	8.9	5.2	5.8
Sandusky, OH	312	8.9	4.5	4.9	5.8
San Francisco-Oakland-Fremont, CA	1295	15.7	8.3	5.7	5.1
San Jose-Sunnyvale-Santa Clara, CA	468	17.8	8.8	5.5	4.9
San Luis Obispo-Paso Robles, CA	42	20.6	6.5	4.3	4.6
Santa Cruz-Watsonville, CA	393	20.8	6.1	4.9	5.1
Santa Fe, NM	38	15.4	6.1	4.8	6.3
Santa Rosa-Petaluma, CA	173	19.4	6.4	4.7	5.4
Savannah, GA	168	13.6	7.2	5.6	5.5
Scranton-Wilkes-Barre, PA	346	10.5	5.2	5.6	5.0
Seattle-Tacoma-Bellevue, WA	326	18.1	8.7	6.3	4.4
Sebastian-Vero Beach, FL	100	17.0	7.6	6.1	4.2
Sheboygan, WI	197	11.1	4.2	9.1	4.6
Sherman-Denison, TX	90	11.4	7.0	3.6	4.9
Shreveport-Bossier City, LA	134	11.6	6.0	5.6	7.4
Sioux City, IA-NE-SD	67	11.6	4.8	5.7	5.8
Sioux Falls, SD	52	14.7	6.8	4.6	5.2
South Bend-Mishawaka, IN-MI	304	11.8	4.8	5.9	7.5
Spartanburg, SC	242	12.4	5.5	4.5	7.2
Spokane, WA	182	15.7	6.5	4.7	4.7
Springfield, IL	159	10.1	4.7	5.0	7.8
Springfield, MA	352	9.7	5.4	6.1	4.1
Springfield, MO	82	15.9	8.5	5.6	4.5
Springfield, OH	379	9.2	4.1	5.1	5.0
State College, PA	102	11.4	5.2	4.4	4.4
Stockton, CA	231	15.0	6.6	4.9	6.1
Sumter, SC	130	11.7	5.3	5.6	7.2
Syracuse, NY	271	10.8	6.0	5.0	5.0
Tallahassee, FL	82	15.7	7.1	4.8	5.6
Tampa-St. Petersburg-Clearwater, FL	588	17.3	9.6	5.5	5.1
Terre Haute, IN	119	10.1	4.0	5.2	7.7
Texarkana, TX-Texarkana, AR	73	11.9	5.2	5.2	7.3
Toledo, OH	404	10.6	5.2	5.5	8.5
Topeka, KS	62	13.2	4.3	4.8	11.3
Trenton-Ewing, NJ	1387	9.7	7.3	4.3	4.5
Tucson, AZ	53	16.6	6.5	6.9	5.4
Tulsa, OK	104	14.3	7.0	5.2	5.5
Tuscaloosa, AL	60	14.3	6.3	7.6	4.8
Tyler, TX	127	12.7	7.3	6.2	6.0
Utica-Rome, NY	125	10.3	4.6	3.9	6.1
Valdosta, GA	56	12.6	5.1	5.0	7.2
Vallejo-Fairfield, CA	249	16.9	6.1	5.5	5.1
Victoria, TX	38	14.7	7.3	5.6	7.8
Vineland-Millville-Bridgeton, NJ	279	10.8	5.3	4.7	7.3
Virginia Beach-Norfolk-Newport News, VA-NC	440	14.4	6.9	4.6	5.3
Visalia-Porterville, CA	47	16.0	6.2	4.2	4.2
Waco, TX	158	12.5	6.0	5.7	5.4
Warner Robins, GA	197	12.7	6.9	3.6	6.0
Washington-Arlington-Alexandria, DC-VA-MD-WV	599	14.0	7.9	5.8	5.5
Waterloo-Cedar Falls, IA	115	12.9	4.7	4.6	5.2
Wausau, WI	69	13.1	4.1	6.6	6.1
Wenatchee-East Wenatchee, WA	14	14.5	7.5	3.6	3.1
Wheeling, WV-OH	197	8.7	4.0	3.6	7.9
Wichita, KS	109	13.6	6.0	5.3	5.6
Wichita Falls, TX	53	11.1	4.7	5.2	7.2
Williamsport, PA	96	9.8	5.1	4.0	9.9
Wilmington, NC	82	15.0	7.9	6.8	4.4
Winchester, VA-WV	62	12.2	5.4	3.6	4.5
Winston-Salem, NC	217	13.8	5.9	4.8	5.2
Worcester, MA	425	9.8	5.8	5.7	4.9
Yakima, WA	38	16.1	5.5	4.7	4.0
York-Hanover, PA	333	9.6	5.6	5.1	6.3
Youngstown-Warren-Boardman, OH-PA	392	10.4	4.6	4.8	7.9
Yuba City, CA	79	16.6	6.1	5.3	5.3
Yuma, AZ	14	13.3	5.4	6.1	7.2
Minimum	3	8.5	3.4	3.2	2.6
25th percentile	73	11.9	5.3	4.8	4.8
Median	134	13.3	6.1	5.2	5.5
75th percentile	239	15.6	7.2	5.7	6.5
Maximum	2480	27.4	12.0	11.6	15.2

Table B.1: List of MSAs included with MSA characteristics

B.1 Migration and population growth

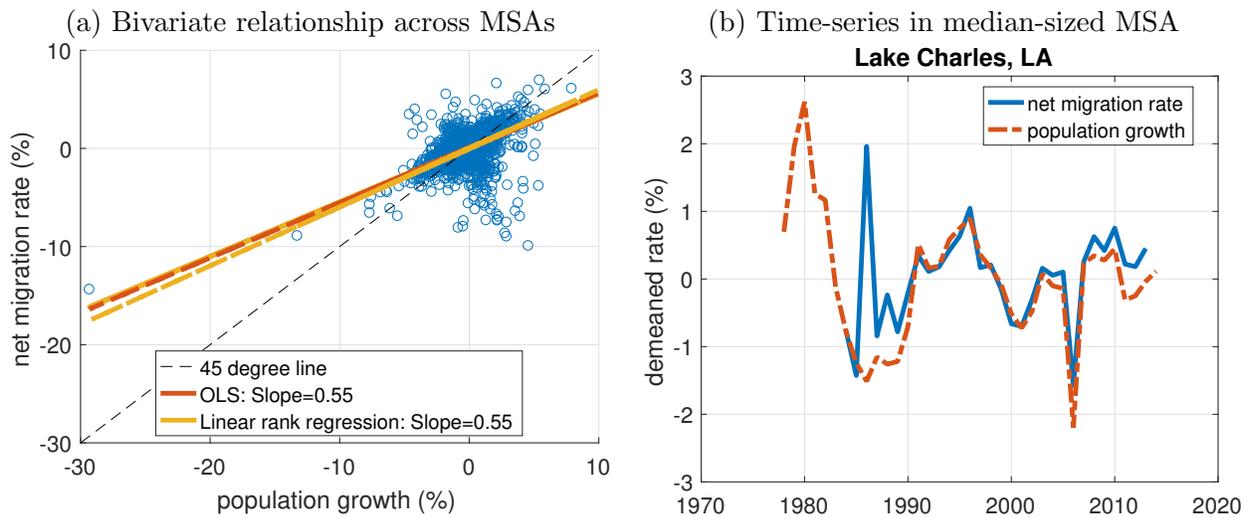
Throughout our discussion, we treat population growth as driven by domestic migration. Figure B.4 contrasts the responses of population growth and the net migration rate to both identified shocks. The migration response is roughly 0.6 times that of population growth and has the same shape for both shocks. We interpret the difference between the responses of the net migration rate and population growth as reflecting measurement error.



The underlying house price data are from CoreLogic Solutions.

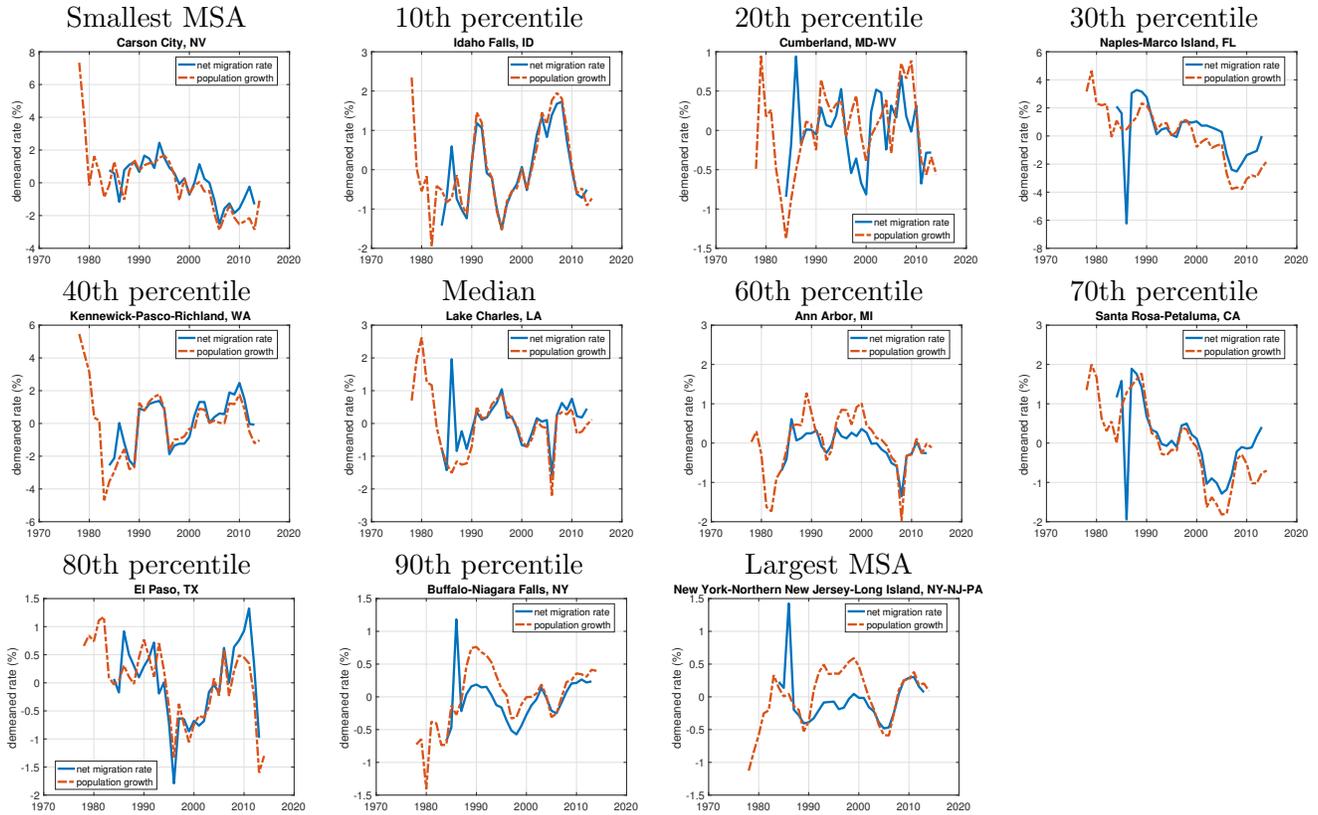
Figure B.4: Impulse-responses of population growth and the net migration rate to the identified shocks: Baseline VAR, 1986 to 2013.

In principle, international migration or changes in fertility and mortality could account for the difference between our measure of migration and population growth. However, Figure B.5(a) shows that if we regress the net migration rate on the population growth rate, after purging it of MSA fixed effects, we find a coefficient of 0.55. This is about the factor of proportionality of the VAR responses.



The figure compares the net migration rate with the population growth rate. Panel (a) shows a scatter plot pooled across time and MSAs, while panel (b) compares the time series for a median-sized MSA (measured in 1986). The figure is consistent with our finding that the migration impulse-responses are a scaled-down version of population growth because of attenuation bias due to measurement error in net migration rates. The scale difference in impulse-responses is close to the slope coefficient in regressions in panel (a). As our example for the MSA with median population size in 1986 in panel (b) shows, the net migration rate often tracks the population growth rate closely, but can, at times, differ erratically, indicating measurement error in migration rates.

Figure B.5: Relationship between net migration rates and population growth rates



For most periods, population growth and the net migration rate track each other in MSAs. Given that the two series move closely together for most periods, we interpret the occasional deviations of the net migration rate from the population growth rate as measurement error. We show the MSAs with the smallest and largest population in 1986, the start of the migration series, and MSAs next to the deciles of the population size distribution.

Figure B.6: Net migration rate and population growth rate for MSAs of various sizes

C Proofs

C.1 Proof of Proposition 2: Identification

Here we re-state and then prove Proposition 2.

Proposition 2 (Identifying shocks.). *Let $\mathbf{V} = \tilde{\mathbf{B}}\tilde{\mathbf{B}}'$ and $\mathbf{\Gamma} = \mathbf{V}'_{n+1:n+n_z, 1:n}$. Partition $\mathbf{\Gamma} = [\mathbf{\Gamma}'_1, \mathbf{\Gamma}'_2]'$, where $\mathbf{\Gamma}_1$ is $n_z \times n_z$. Assume $\mathbf{\Gamma}_1$ is invertible, so that $\boldsymbol{\kappa} = \mathbf{\Gamma}_2\mathbf{\Gamma}_1^{-1}$ is well defined.*

(a) *If the first instrument is correlated only with the first shock, we have $\boldsymbol{\beta}_{[1]} = [1, (\mathbf{\Gamma}_2\mathbf{e}_1)' \frac{1}{\mathbf{\Gamma}_1\mathbf{e}_1}]' \times \bar{c} \propto \mathbf{\Gamma}\mathbf{e}_1$ for $\bar{c} > 0$ (defined in the proof).*

(b) *We can factor $(\mathbf{S}_1\mathbf{S}'_1)^{1/2} = [\boldsymbol{\nu}_1, \boldsymbol{\nu}_2]$ in (15), where $\boldsymbol{\nu}_1 = \bar{c}_1(\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})\mathbf{\Gamma}_1\mathbf{e}_1$ and $\boldsymbol{\nu}_2 = \mathbf{F}\text{chol}(\mathbf{\Lambda})$. Here, \mathbf{F} are the $n_z - 1$ eigenvectors and $\mathbf{\Lambda}$ the diagonal matrix of strictly positive eigenvalues of $\mathbf{S}_1\mathbf{S}'_1 - \boldsymbol{\nu}_1\boldsymbol{\nu}'_1$. Then the first identified shock is identified only from the first instrument, i.e., $\boldsymbol{\beta}_{[1]}\mathbf{e}_1 \propto \mathbf{\Gamma}\mathbf{e}_1$.*

Proof. Part (a): The treatment in [Stock and Watson \(2012\)](#) shows most clearly that with a single shock, the impact response is proportional to $\mathbf{\Gamma}$. But even if it requires more algebra, we now also show how this follows from (15). Here, we follow the notation in [Drautzburg \(forthcoming, Appendix A\)](#), except for substituting $\boldsymbol{\beta}$ for $\boldsymbol{\alpha}$.

To identify a single shock from (15), set $n_z = 1$. $\boldsymbol{\eta} = \boldsymbol{\beta}_{12}\boldsymbol{\beta}_{22}^{-1}$ and $\boldsymbol{\kappa} = \boldsymbol{\beta}_{21}\boldsymbol{\beta}_{11}^{-1} = \mathbf{\Gamma}_2\mathbf{\Gamma}_1^{-1}$. It follows that $\mathbf{S}_1 = (\boldsymbol{\beta}_{11} - \boldsymbol{\beta}_{12}\boldsymbol{\beta}_{22}^{-1}\boldsymbol{\beta}_{21})$. By construction, $\boldsymbol{\beta}_{11}$ is the conditional standard deviation of the first variable attributable to the identified shock: $\boldsymbol{\beta}_{11} = \sqrt{\boldsymbol{\Sigma}_{11} - f(\boldsymbol{\Sigma}, \boldsymbol{\kappa})}$, normalizing the sign of the shock so that the impact-response is positive. $\boldsymbol{\kappa} = \mathbf{\Gamma}_2\mathbf{\Gamma}_1^{-1}$, and $f(\boldsymbol{\Sigma}, \boldsymbol{\kappa}) \equiv \boldsymbol{\beta}_{12}\boldsymbol{\beta}'_{12} = (\boldsymbol{\Sigma}'_{12} - \boldsymbol{\kappa}\boldsymbol{\Sigma}_{11})'(\mathbf{Z}\mathbf{Z}')^{-1}(\boldsymbol{\Sigma}'_{12} - \boldsymbol{\kappa}\boldsymbol{\Sigma}_{11})$ with $\mathbf{Z}\mathbf{Z}' = \begin{bmatrix} \boldsymbol{\kappa} & -\mathbf{I}_{m-m_z} \end{bmatrix} \boldsymbol{\Sigma} \begin{bmatrix} \boldsymbol{\kappa}' \\ -\mathbf{I}_{m-m_z} \end{bmatrix}$ ([Drautzburg, forthcoming, Appendix A](#)).

To prove that $\boldsymbol{\beta}_{21} = \mathbf{\Gamma}_2$ is as desired, use the Woodbury matrix identity to write

$$\begin{aligned} (\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\eta})^{-1} &= \mathbf{I} + \boldsymbol{\kappa}(\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1}\boldsymbol{\eta} \\ &= \mathbf{I} + \boldsymbol{\kappa}\boldsymbol{\eta}(\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1}, \end{aligned}$$

where the second equality uses that $(\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1}$ is a scalar with $n_z = 1$.

Consequently:

$$\begin{aligned} \boldsymbol{\beta}_{21} &= (\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\eta})^{-1}\boldsymbol{\kappa}\mathbf{S}_1 \\ &= \boldsymbol{\kappa}\mathbf{S}_1 + \boldsymbol{\kappa}\boldsymbol{\eta}\boldsymbol{\kappa}(\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1}\mathbf{S}_1 \\ &= \boldsymbol{\kappa}\mathbf{S}_1 + \boldsymbol{\kappa}(\boldsymbol{\eta}\boldsymbol{\kappa} - \mathbf{I} + \mathbf{I})(\mathbf{I} - \boldsymbol{\beta}\boldsymbol{\kappa})^{-1}\mathbf{S}_1 \\ &= \boldsymbol{\kappa}\mathbf{S}_1 - \boldsymbol{\kappa}(\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})(\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1}\mathbf{S}_1 + \boldsymbol{\kappa}(\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1}\mathbf{S}_1 \end{aligned}$$

$$= \boldsymbol{\kappa}(\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1}\mathbf{S}_1 \equiv \boldsymbol{\kappa}\boldsymbol{\beta}_{11} \equiv \boldsymbol{\Gamma}_2 \frac{\boldsymbol{\beta}_{11}}{\boldsymbol{\Gamma}_1}. \quad (\text{C.1})$$

Therefore, $\boldsymbol{\beta}_{[1]} = [\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2]' \frac{\boldsymbol{\beta}_{11}}{\boldsymbol{\Gamma}_1}$.

In Proposition 2 we consider $n_z > 1$, so that we need to replace the scalar $\boldsymbol{\Gamma}_1$ here with $[\boldsymbol{\Gamma}_1]_{11} = \mathbf{e}'_1 \boldsymbol{\Gamma}_1 \mathbf{e}_1$. Consequently, the constant in Proposition 2 is given by $\bar{c} = \frac{\boldsymbol{\beta}_{11}}{[\boldsymbol{\Gamma}_1]_{11}}$ where $\boldsymbol{\beta}_{11} = \sqrt{\boldsymbol{\Sigma}_{11} - f(\boldsymbol{\Sigma}, \boldsymbol{\kappa})}$, normalizing the sign of the shock so that the impact-response is positive.

Part (b): We proceed in two parts. First, we prove that $[\boldsymbol{\nu}_1, \boldsymbol{\nu}_2][\boldsymbol{\nu}_1, \boldsymbol{\nu}_2]' = \mathbf{S}_1 \mathbf{S}'_1$. Second, we prove that for $(\mathbf{S}_1 \mathbf{S}'_1)^{1/2} = [\boldsymbol{\nu}, \boldsymbol{\nu}_2]$ it holds that $\boldsymbol{\beta}_{[1]} \mathbf{e}_1 \propto \boldsymbol{\Gamma} \mathbf{e}_1$.

(1) Note that the $n_z \times n_z$ matrix $\mathbf{S}_1 \mathbf{S}'_1 - \boldsymbol{\nu}_1 \boldsymbol{\nu}'_1$ is symmetric of rank $n_z - 1$. Therefore:

$$\mathbf{S}_1 \mathbf{S}'_1 - \boldsymbol{\nu}_1 \boldsymbol{\nu}'_1 = [\mathbf{F} \quad \mathbf{f}_\perp] \begin{bmatrix} \boldsymbol{\Lambda} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{F}'_1 \\ \mathbf{f}'_\perp \end{bmatrix} = \mathbf{F} \boldsymbol{\Lambda} \mathbf{F}',$$

where \mathbf{F} are $n_z - 1$ normed $n_z \times 1$ eigenvectors associated with the strictly positive eigenvalue λ_i , where $\boldsymbol{\Lambda} = \text{diag}([\lambda_i]_{i=1}^{n_z-1})$. Because the eigenvectors are normed $\mathbf{F}'\mathbf{F} = \mathbf{I}_{n_z-1}$, $\mathbf{f}_\perp' \mathbf{f}_\perp = 1$ and $\mathbf{F}'\mathbf{f}_\perp = 0$. Therefore $\boldsymbol{\nu}_2 = \mathbf{F} \text{chol}(\boldsymbol{\Lambda})$ satisfies $\boldsymbol{\nu}_2 \boldsymbol{\nu}'_2 = \mathbf{S}_1 \mathbf{S}'_1 - \boldsymbol{\nu}_1 \boldsymbol{\nu}'_1$ as desired.

(2) Rewrite $\boldsymbol{\beta}_{[1]} \mathbf{e}_1$ from (15):

$$\boldsymbol{\beta}_{[1]} \mathbf{e}_1 = \begin{bmatrix} (\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1} \\ (\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\eta})^{-1} \boldsymbol{\kappa} \end{bmatrix} \boldsymbol{\nu}_1.$$

By construction:

$$(\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1} \boldsymbol{\nu}_1 = (\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa})^{-1} \bar{c}_1 (\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa}) \boldsymbol{\Gamma}_1 \mathbf{e}_1 = \bar{c}_1 \boldsymbol{\Gamma}_1 \mathbf{e}_1, \quad (\text{C.2})$$

as desired.

It remains to show that

$$(\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\eta})^{-1} \boldsymbol{\kappa} \boldsymbol{\nu}_1 = \bar{c}_1 \boldsymbol{\Gamma}_2 \mathbf{e}_1 \quad (\text{C.3})$$

to ensure the same factor of proportionality. Plugging in for $\boldsymbol{\nu}_1$:

$$\begin{aligned} (\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\beta})^{-1} \boldsymbol{\kappa} \bar{c}_1 (\mathbf{I} - \boldsymbol{\eta}\boldsymbol{\kappa}) \boldsymbol{\Gamma}_1 \mathbf{e}_1 &= \bar{c}_1 \boldsymbol{\Gamma}_2 \mathbf{e}_1. \\ \Leftrightarrow \bar{c}_1 (\boldsymbol{\kappa} - \boldsymbol{\kappa}\boldsymbol{\eta}\boldsymbol{\kappa}) \boldsymbol{\Gamma}_1 \mathbf{e}_1 &= \bar{c}_1 (\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\beta}) \boldsymbol{\Gamma}_2 \mathbf{e}_1. \\ \Leftrightarrow \bar{c}_1 (\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\eta}) \boldsymbol{\kappa} \boldsymbol{\Gamma}_1 \mathbf{e}_1 &= \bar{c}_1 (\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\eta}) \boldsymbol{\Gamma}_2 \mathbf{e}_1. \\ \Leftrightarrow \bar{c}_1 (\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\eta}) \boldsymbol{\Gamma}_2 \mathbf{e}_1 &= \bar{c}_1 (\mathbf{I} - \boldsymbol{\kappa}\boldsymbol{\eta}) \boldsymbol{\Gamma}_2 \mathbf{e}_1. \end{aligned}$$

The second to last equality uses that $\boldsymbol{\kappa} = \boldsymbol{\Gamma}_2 \boldsymbol{\Gamma}_1^{-1}$. Combining (C.2) and (C.3) it follows that $\boldsymbol{\beta}_{[1]} \mathbf{e}_1 \propto$

$\Gamma \mathbf{e}_1$.

□

C.2 Stacking and vectorizing the VAR

As it stands, equation (10) is part of a system of simultaneous equations and as such untractable. To simplify, it is useful to bring it into vector notation. As a first step, rewrite equation (10):

$$\mathbf{u}_{mt} = \mathbf{R} [\mathbf{u}_{1t} \ \dots \ \mathbf{u}_{Nt}] \mathbf{D}' \mathbf{e}_{m,M} + \mathbf{B} \boldsymbol{\varepsilon}_{mt} \equiv \mathbf{R} \mathbf{U}_t \mathbf{D}' \mathbf{e}_{m,M} + \mathbf{B} \boldsymbol{\varepsilon}_{mt},$$

where $\mathbf{e}_{m,M}$ is an $M \times 1$ selection vector of zeros except for a one in its m th position. When obvious from the context, we drop the second subscript of the selection vector in what follows.

Stacking the model horizontally for each period across all MSAs:

$$\mathbf{Y}_t \equiv \begin{bmatrix} \mathbf{y}'_{1t} \\ \vdots \\ \mathbf{y}'_{Nt} \end{bmatrix}' = \mathbf{A} \mathbf{X}_{t-1} + \boldsymbol{\mu} + \boldsymbol{\eta}_t \mathbf{1}'_M + \mathbf{u}_t, \quad \mathbf{A} \equiv [\mathbf{A}_1 \ \dots \ \mathbf{A}_k], \mathbf{X}_{t-1} \equiv \begin{bmatrix} \mathbf{Y}_{t-1} \\ \vdots \\ \mathbf{Y}_{t-k-1} \end{bmatrix} \quad (\text{C.4})$$

$$\mathbf{U}_t \equiv \begin{bmatrix} \mathbf{u}'_{1,t} \\ \vdots \\ \mathbf{u}'_{M,t} \end{bmatrix}' = \mathbf{R} \mathbf{U}_t \mathbf{D}' + \mathbf{B} \boldsymbol{\varepsilon}_t. \quad (\text{C.5})$$

Using the vec operator rule that $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \text{vec}(\mathbf{B})$ and the special case of $\text{vec}(\mathbf{BC}) = (\mathbf{C}' \otimes \mathbf{I}) \text{vec}(\mathbf{B})$, this can in turn be rewritten as:

$$\text{vec}(\mathbf{Y}_t) \equiv (\mathbf{I}_M \otimes \mathbf{A}) \text{vec}(\mathbf{X}_{t-1}) + \text{vec}(\mathbf{U}_t), \quad (\text{12})$$

$$\begin{aligned} \text{vec}(\mathbf{U}_t) &= (\mathbf{D} \otimes \mathbf{R}) \text{vec}(\mathbf{U}_t) + (\mathbf{I}_M \otimes \mathbf{B}) \text{vec}(\boldsymbol{\varepsilon}_t) \\ &= (\mathbf{I}_{Nq} - (\mathbf{D} \otimes \mathbf{R}))^{-1} ((\mathbf{I}_M \otimes \mathbf{B}) \text{vec}(\boldsymbol{\varepsilon}_t)), \end{aligned} \quad (\text{13})$$

where $\text{vec}(\boldsymbol{\varepsilon}_t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{Nq})$. This form of the spatial VAR is tractable as it expresses the forecast error in terms of the *iid* standard normal residuals $\text{vec}(\boldsymbol{\varepsilon}_t)$ and therefore allows us to write down the likelihood function or to derive the form of the impulse-responses.

C.3 Rewriting the likelihood function

Lemma 1. *Let $\tilde{\mathbf{R}}$ be a $(n + n_z + n_p) \times (n + n_z + n_p)$ diagonal matrix: $\tilde{\mathbf{R}} = \text{diag}([\tilde{\rho}_s])$. Then $|\mathbf{I}_{M(n+n_z+n_p)} - \mathbf{D} \otimes \tilde{\mathbf{R}}| = \prod_{s=1}^{n+n_z+n_p} |\mathbf{I}_M - \tilde{\rho}_s \mathbf{D}|$.*

Proof. To derive the equality, first define a square $M(n + n_z + n_p) \times M(n + n_z + n_p)$ commutation matrix $\mathbf{P}_{M(n+n_z+n_p)}$. $|\mathbf{P}_{M(n+n_z+n_p)}| = |\mathbf{P}_{(n+n_z+n_p)M}| = \pm 1$, depending on whether M and $n + n_z + n_p$ are even or

odd, see [Lütkepohl \(2005, p. 664, results \(12\) and \(24\)\)](#) for this and the following result: $\mathbf{P}_{M(n+n_z+n_p)}\mathbf{D} \otimes \tilde{\mathbf{R}}\mathbf{P}_{(n+n_z+n_p)M} = \mathbf{P}_{M(n+n_z+n_p)}\mathbf{D} \otimes \tilde{\mathbf{R}}\mathbf{P}_{M(n+n_z+n_p)}^{-1} = \tilde{\mathbf{R}} \otimes \mathbf{D}$.

Putting these results from [Lütkepohl \(2005\)](#) together yields:

$$\begin{aligned} |\mathbf{I}_{M(n+n_z+n_p)} - \mathbf{D} \otimes \tilde{\mathbf{R}}| &= |\mathbf{P}_{M(n+n_z+n_p)}| |\mathbf{I}_{M(n+n_z+n_p)} - \mathbf{D} \otimes \tilde{\mathbf{R}}| |\mathbf{P}_{M(n+n_z+n_p)}^{-1}| \\ &= |\mathbf{I}_{M(n+n_z+n_p)} - \mathbf{P}_{M(n+n_z+n_p)}\mathbf{D} \otimes \tilde{\mathbf{R}}\mathbf{P}_{M(n+n_z+n_p)}^{-1}| \\ &= |\mathbf{I}_{M(n+n_z+n_p)} - \tilde{\mathbf{R}} \otimes \mathbf{D}|. \end{aligned}$$

Because $\tilde{\mathbf{R}}$ is diagonal, the matrix inside the determinant is block-diagonal. Using the rule for the determinant of partitioned matrices ([Lütkepohl, 2005, p. 660](#)) repeatedly gives us that $|\mathbf{I}_{M(pq+n_z+n_p)} - \tilde{\mathbf{R}} \otimes \mathbf{D}| = \prod_{s=1}^{n+n_z+n_p} |\mathbf{I}_M - \tilde{\rho}_s \mathbf{D}|$.

Last, $|\mathbf{I}_{M(n+n_z+n_p)} - \mathbf{D} \otimes \tilde{\mathbf{R}}| = |\mathbf{I}_{M(n+n_z+n_p)} - \tilde{\mathbf{R}} \otimes \mathbf{D}|$, so that the desired result follows. \square

D Heterogeneous coefficients

We also examine whether the dynamics differ depending on MSA characteristics. To do so, we split the sample based on MSA characteristics. When possible, we choose characteristics before the start of our estimation sample, such as the population density in 1976 or the startup entry rate in 1978. We then estimate VAR-coefficients, within-MSA VAR-covariances, and spatial correlations with neighboring MSAs separately for each group. However, we do take account of the fact that the errors are dependent across groups.³⁴

Specifically, a simple transformation still purges the overall MSA-specific error term of their spatial dependence. Take the i th component of the overall error term in (13). Let $G(m)$ map MSA m to group g . Define $\mathbf{V}_t^i = \mathbf{e}_i' [\mathbf{B}_{G(m)} \boldsymbol{\varepsilon}_{m,t}]_{m=1}^M$ to be the vector of within-MSA forecast errors. With heterogeneous spatial correlations, the overall error term becomes:

$$\begin{aligned} \mathbf{u}_t^i &= \left[\sum_n d_{mn} \rho_{G(m)}^i u_{n,t}^i + \mathbf{v}_{m,t}^i \right]_{m=1}^M \\ &= \mathbf{u}_t^i \mathbf{D} \text{diag}([\rho_{G(m)}^i]_m) + \mathbf{v}_t^i \\ &= \mathbf{v}_t^i (\mathbf{I} - \mathbf{D} \text{diag}([\rho_{G(m)}^i]_m))^{-1}. \end{aligned} \quad (\text{D.1})$$

Post-multiplying the spatial errors by $\mathbf{I} - \mathbf{D} \text{diag}([\rho_{G(m)}^i]_m)$ therefore rids the error terms of their spatial dependence. We can estimate $\boldsymbol{\Sigma}_g = \mathbf{B}_g \mathbf{B}_g'$ from the transformed error terms of all MSAs m that belong to group g .

Thus we conduct our inference separately across groups of MSAs, except for estimating the spatial correlation coefficients. The group-specific spatial correlation coefficients still maximize the joint quasi-likelihood across MSAs of all groups. Of course, the bootstrap re-samples from v_t^i , using either an *iid* or a block bootstrap, and then re-introduces the estimated spatial correlation by post-multiplying with $(\mathbf{I} - \mathbf{D} \text{diag}([\rho_{G(m)}^i]_m))^{-1}$. This, of course, affects the quasi-likelihood.

It is useful to rewrite the quasi-likelihood function by commuting rows and columns of the spatial transform. This corresponds to re-ordering the vector of VAR-residuals. It simplifies performing the computations and extending the model below. Since, by Lemma 1, $|\mathbf{I}_{M(n+n_z+n_p)} - \mathbf{D} \otimes \tilde{\mathbf{R}}| = \prod_{s=1}^{n+n_z+n_p} |\mathbf{I}_M - \tilde{\rho}_s \mathbf{D}|$, it follows that:

$$\log L^c = -\frac{1}{2} \left(MT \log(2\pi) + 2T \sum_{s=1}^{n+n_z+n_p} \log(|\mathbf{I}_M - \tilde{\rho}_s \mathbf{D}|) - MT \log(|\hat{\mathbf{V}}|) \right) + c. \quad (\text{D.2})$$

³⁴This is similar in spirit to [Bonhomme and Manresa \(2015\)](#) with group fixed effects and heterogeneous coefficients, but with known group assignments.

Now it is easy to see how the likelihood in (D.2) changes when the spatial transform becomes group-specific:

$$\log L^c = -\frac{1}{2} \left(MT \log(2\pi) + 2T \sum_{s=1}^{n+n_z} \log(\mathbf{I} - \mathbf{D} \text{diag}([\rho_{G(m)}^s]_m)) - MT \log(|\hat{\mathbf{V}}|) \right) + c. \quad (\text{D.3})$$

E TFP calculation

We use the approach of [Olley and Pakes \(1996\)](#) to estimate TFP growth, following [Imrohoroglu and Tüzel \(2014\)](#) for the implementation. All data are Compustat data from [S&P Global Market Intelligence \(2018\)](#) via Wharton Research Data Services (WRDS). Specifically, we calculate TFP growth as follows:

- Define variables for OP regression
 1. Calculate output
 - Sales minus operating income before depreciation OIBDP labor expenditure (EMP times average wage)
 - Divide by the GDP deflator and take the log
 2. Calculate capital
 - Calculate average age (DPACT (Accumulated Depreciation and Amortization)/DP (Depreciation and Amortization))
 - Take 3-year moving average of average age
 - Calculate average year (Birth + Average Age)
 - Calculate capital as PPEGT (Gross Property, Plant and Equipment) multiplied by the ratio of the current investment deflator to the investment deflator of the average year
 - Take the log of computed capital and lag it one year
 3. Calculate employment: log of employment
 4. Calculate investment: log of CAPXV (Capital Expenditures) divided by the investment price deflator
- Calculate converted industry codes
 - If in any given year and any given SIC-3 code there are less than 20 (10) firm observations, then collapse the SIC-3 code with the corresponding SIC-2 code
 - If in any given year and any given SIC-2 code there are less than 20 (10) firm observations, then collapse the SIC-1 code with the corresponding SIC-1 code
- By converted industry j : [Olley and Pakes \(1996\)](#) Regression for TFP levels
 - Regress output on employment and second order polynomial of investment, age, and capital

- Estimate beta coefficient for employment
- Calculate ϕ_{it} as predicted sale minus employment times $\hat{\beta}_e^{(j)}$
- Calculate $h_{it} = \phi_{it} - \text{age}_{it}\hat{\beta}_a^{(j)} - k_{t-1}\beta_k$
- Estimate the predicted survival probability \hat{p}_{it} using a logistic regression of survival on second order polynomial for investment, age and capital
- Regress sale minus employment times the employment coefficient on capital, age, and the lagged second order polynomial of h_{it} and \hat{p}_{it}
- Estimate beta coefficients for capital, age and the constant
- Calculate TFP: $TFP_{it} = \text{sale}_{it} - \text{emp}_{it}\hat{\beta}_e^{(j)} - \text{age}_{it}\hat{\beta}_a^{(j)} - k_{i,t-1}\hat{\beta}_k^{(j)} - \hat{\beta}_0^{(j)}$, where j denotes the industry firm i belongs to
- Collapsed firm-level TFP growth $\Delta \ln TFP_{it}$ by year t and converted industry codes j

Requiring ≥ 20 obs	Requiring ≥ 10 obs	
SIC 0	SIC 0	
SIC 1, nec	SIC 1, nec	SIC 5, nec
SIC 2, nec	SIC 10, nec	SIC 50, nec
SIC 3, nec	SIC 131	SIC 51, nec
SIC 4, nec	SIC 138	SIC 514
SIC 5, nec	SIC 2, nec	SIC 541
SIC 6, nec	SIC 20, nec	SIC 581
SIC 7, nec	SIC 201	SIC 59
SIC 8, nec	SIC 23	SIC 6, nec
SIC 20, nec	SIC 24	SIC 65
SIC 26, nec	SIC 25	SIC 679
SIC 28, nec	SIC 26	SIC 7, nec
SIC 30, nec	SIC 27	SIC 701
SIC 33, nec	SIC 28, nec	SIC 73, nec
SIC 34, nec	SIC 282	SIC 737
SIC 35, nec	SIC 283	SIC 738
SIC 36, nec	SIC 284	SIC 799
SIC 45, nec	SIC 289	SIC 8, nec
SIC 48, nec	SIC 291	SIC 80
SIC 49, nec	SIC 3, nec	SIC 87
SIC 50, nec	SIC 308	
SIC 51, nec	SIC 32	
SIC 65, nec	SIC 33, nec	
SIC 73, nec	SIC 331	
SIC 79, nec	SIC 34	
SIC 80, nec	SIC 35, nec	
SIC 87, nec	SIC 353	
SIC 131	SIC 355	
SIC 138	SIC 356	
SIC 283	SIC 357	
SIC 356	SIC 358	
SIC 357	SIC 36, nec	
SIC 366	SIC 362	
SIC 367	SIC 366	
SIC 371	SIC 367	
SIC 382	SIC 369	
SIC 384	SIC 37, nec	
SIC 421	SIC 371	
SIC 481	SIC 381	
SIC 491	SIC 382	
SIC 492	SIC 384	
SIC 493	SIC 39	
SIC 581	SIC 4, nec	
SIC 737	SIC 421	
	SIC 451	
	SIC 481	
	SIC 483	
	SIC 484	
	SIC 49	
	SIC 491	
	SIC 492	
	SIC 493	

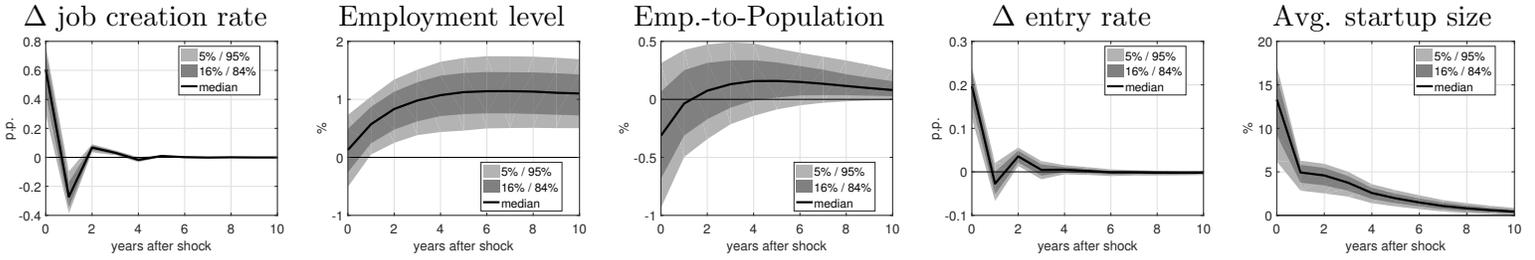
Table E.1: List of industries used for TFP Bartik instrument as a function of the minimum number of firm-years in each industry.

F Additional estimates

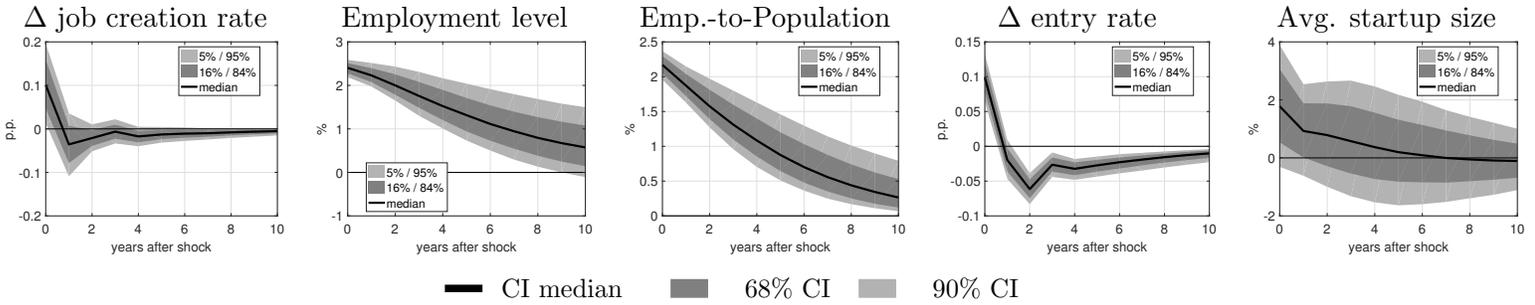
Section F.1 presents additional results for our baseline model. Section F.2 presents various alternative specifications for our baseline model. Section F.3 shows additional results when we split MSAs by initial density. Section F.4 shows all results when we split MSAs by initial firm entry.

F.1 Additional baseline results

(a) Startup shock



(b) Overall labor demand shock



Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. Startup shocks have small but persistent effects on local employment, mostly due to the extensive margin.

Figure F.1: Additional impulse-responses to startup shocks and overall labor demand shocks in our baseline VAR.

Table F.1: Spatial autocorrelation: Bootstrapped likelihood ratio test for differential spatial autocorrelation.

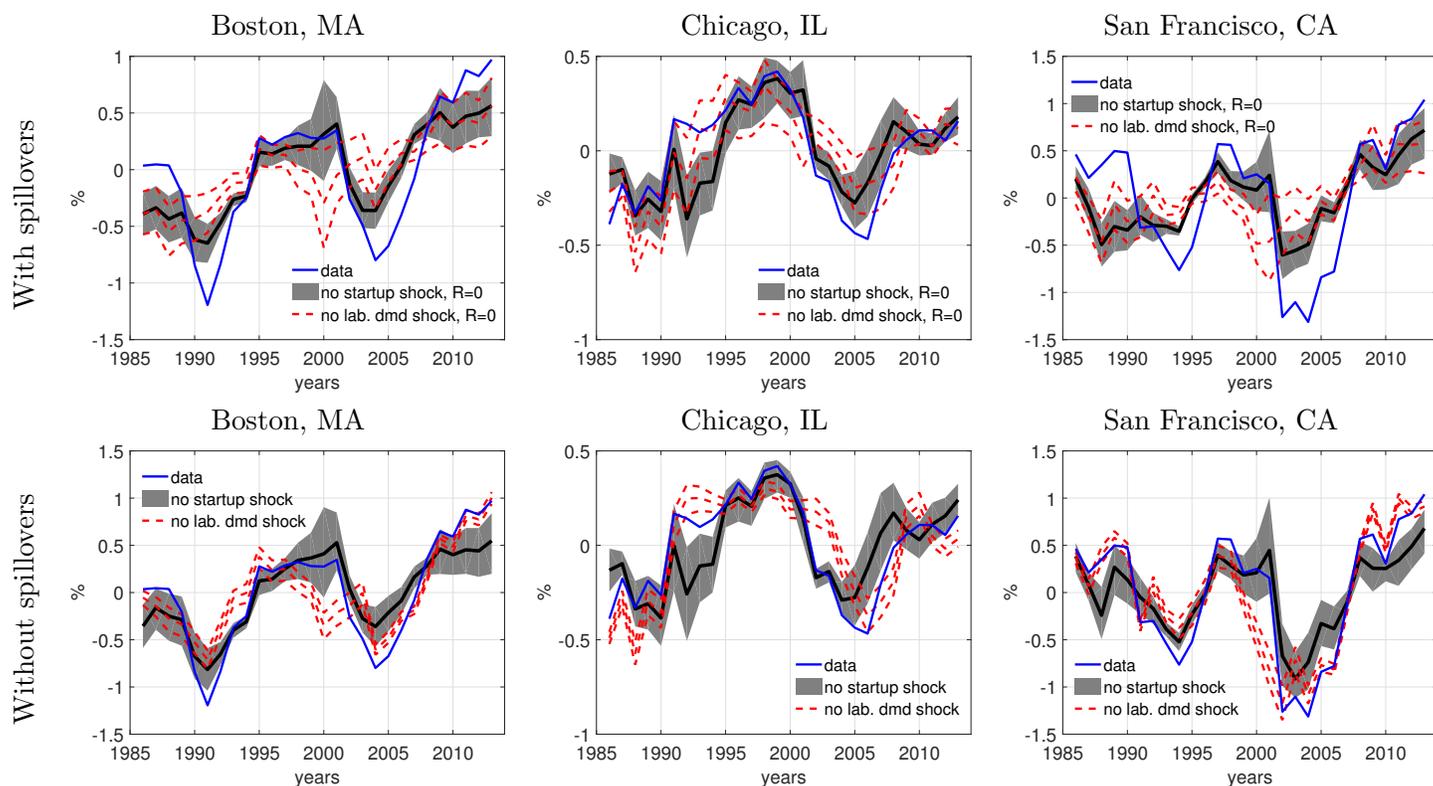
Comparison	Point estimate	Confidence interval				
		1%	5%	Median	95%	99%
H0: Constant spatial correlation						
Same ρ vs. varying ρ s	5.333	0.012	0.017	0.037	0.077	0.107
H0: No spatial correlation						
No spatial correlation vs. varying ρ s	32.307	0.016	0.021	0.046	0.092	0.114

The table shows likelihood ratio test statistic $-2 \ln(L_{restricted}/L_{varying})$, divided by N for legibility with bootstrapped confidence intervals. The hypotheses of a common spatial correlation or no spatial correlation are rejected at the 1% level according to the simulated distribution of the test statistic in this table. Underlying house price data are from CoreLogic Solutions.

Table F.2: VS funding and startup shocks: 2000 to 2013.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VC growth	0.088*	0.559	0.554*	0.346*	0.041***	0.202***	0.183***	0.185***
	(1.96)	(1.65)	(2.07)	(1.90)	(5.03)	(3.30)	(3.37)	(3.29)
Lagged VC growth	0.044	0.191			0.021	0.077		
	(0.79)	(0.65)			(0.83)	(0.67)		
Observations	1295	1295	1726	1767	1284	1284	1708	1750
MSAs	164	164	209	250	163	163	208	250
Years	12	12	13	13	12	12	13	13
Transformation	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Weights	No	No	No	No	Yes	Yes	Yes	Yes

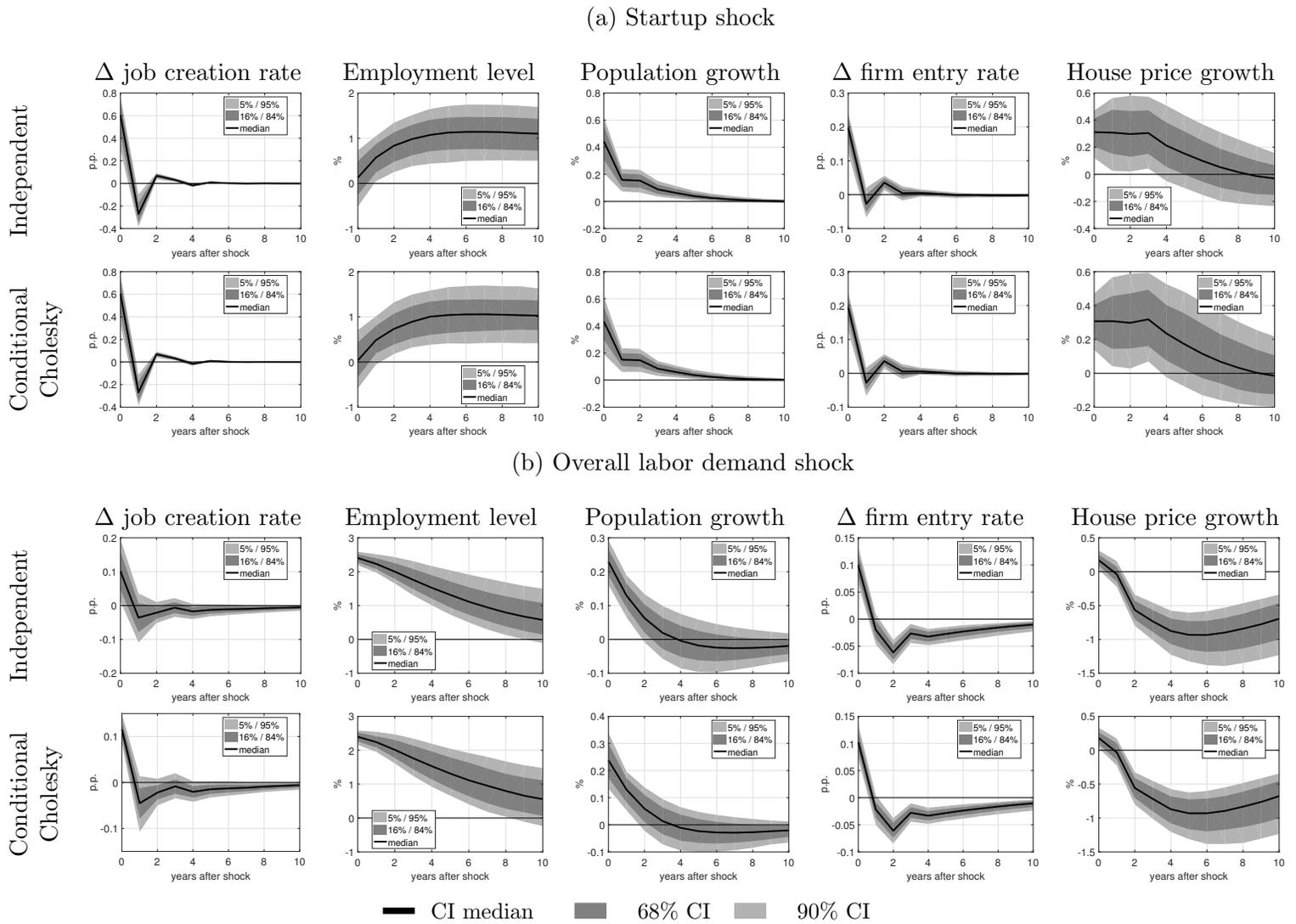
The table shows slope coefficients (t -statistics) for regressions of the median estimated shock realization in the MSAs on the growth in the number of venture capital funded firms. Transformation, if indicated, refers to the logit transform of the VC growth rate: Δvc_t by $\frac{\exp(\Delta vc_t)}{1+\exp(\Delta vc_t)}$. Fixed effects, if indicated, are MSA and year fixed effects. Weights are Stata's `rreg` weights, i.e., combining Cook's distance with downweighting of observations with large residuals. t -statistics in parentheses. Standard errors clustered by MSA and year. * $p < 0.10$ ** $p < .05$ *** $p < .01$. p -values based on the t -distribution with $\#years-1$ degrees of freedom.



This plot quantifies the importance of startup shocks and overall labor demand shocks for three MSAs while turning off the spatial correlation. Turning off the spatial correlation brings the counterfactuals closer to zero, highlighting the role of spatial correlation in explaining outcomes. All variables in deviations net of MSA and year fixed effects. We show 68% confidence intervals.

Figure F.2: Historical contributions to population growth: Comparison of three large MSAs, 1986–2013. The role of spillovers ($\tilde{R} = 0$).

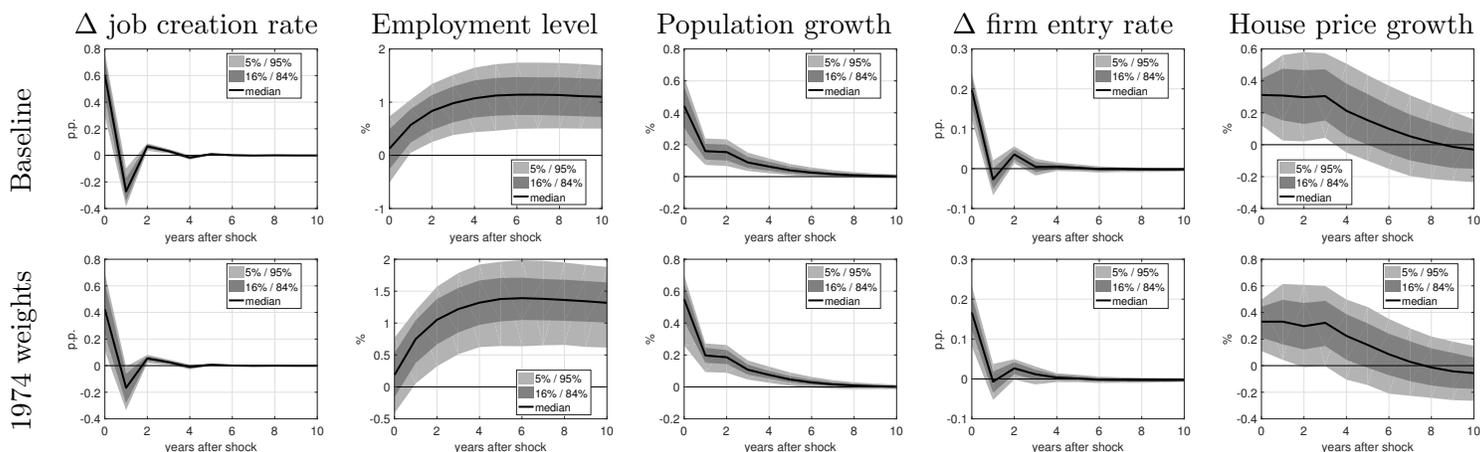
F.2 VAR specification



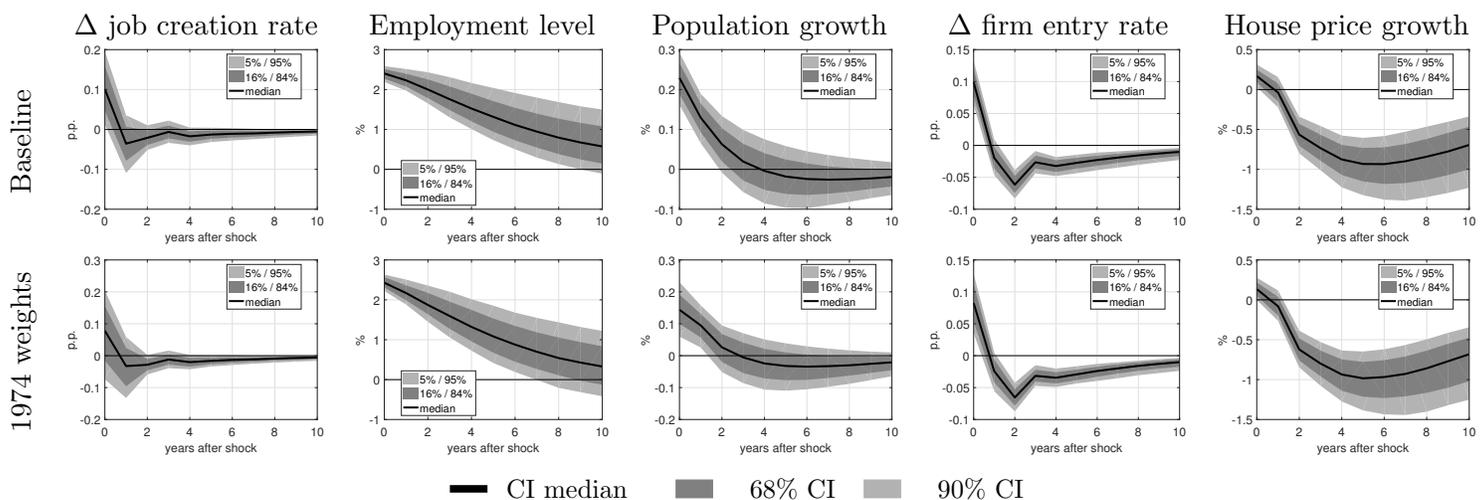
Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. The underlying house price data are from CoreLogic Solutions. We compare two ways to factor the two identified shocks: In our baseline (“independent”), we attribute all the variation in the standard [Bartik \(1991\)](#) instrument to the overall labor demand shock. In the alternative, we choose a Cholesky factorization of the variance attributable to the two identified shocks that orders the overall labor demand shock first (“conditional Cholesky”). Both give almost identical answers.

Figure F.3: Impulse-responses in baseline VAR: Treating the standard Bartik IV as independent vs. ordering it first in a conditional Cholesky factorization.

(a) Startup shock



(b) Overall labor demand shock

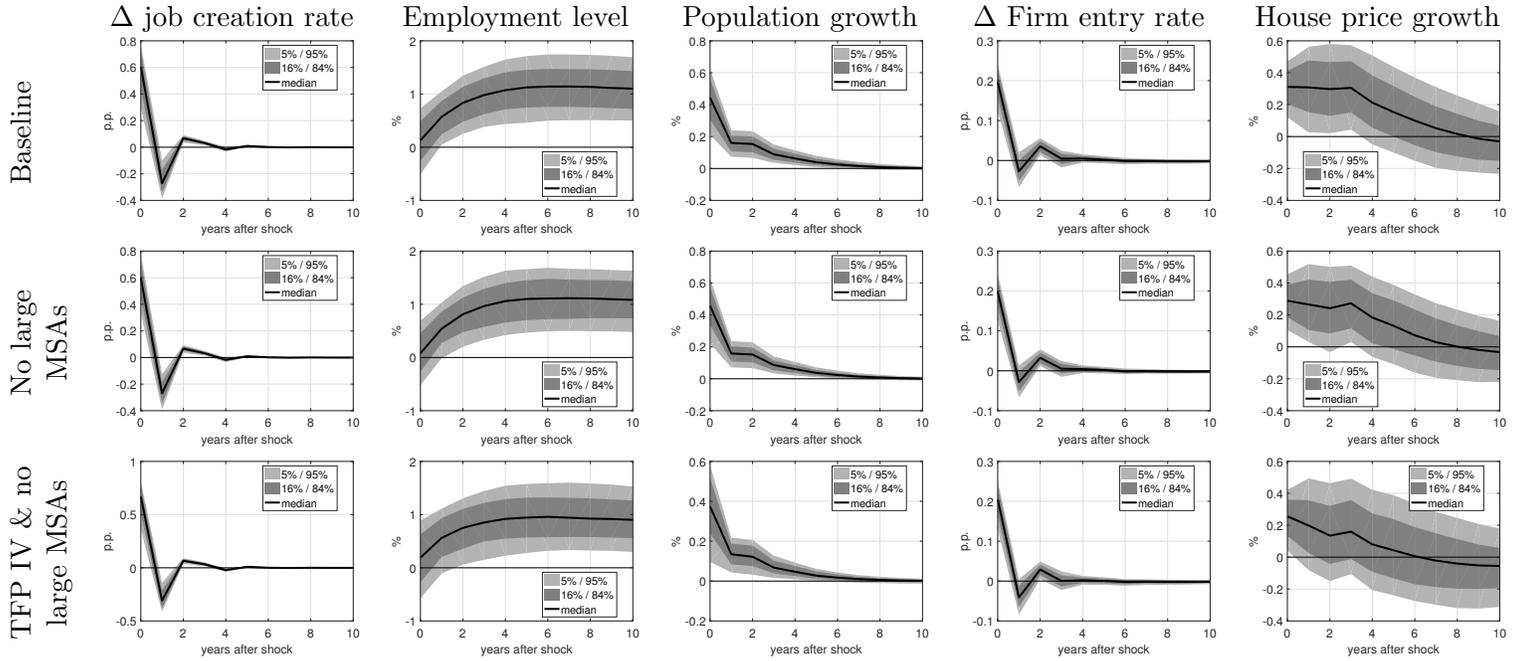


— CI median ■ 68% CI ■ 90% CI

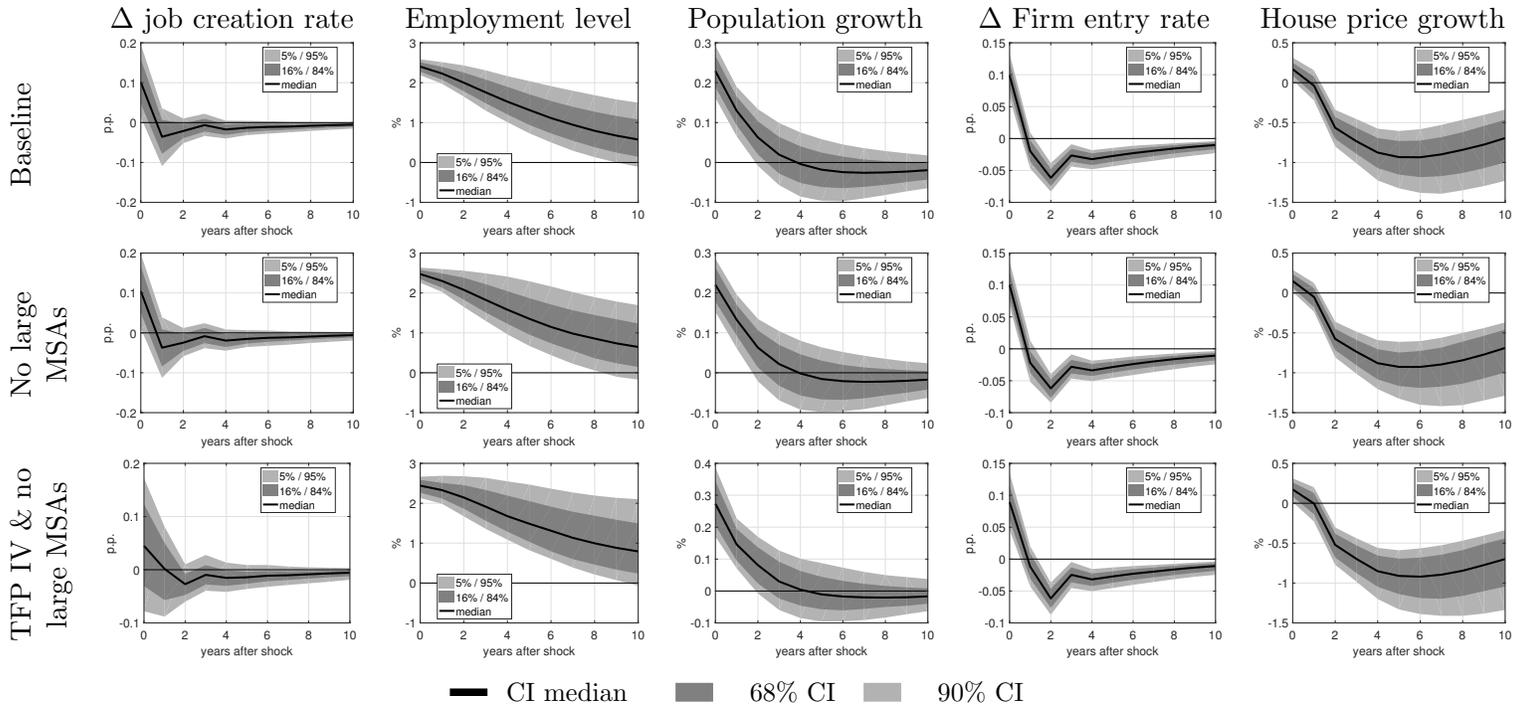
Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. The underlying house price data are from CoreLogic Solutions. We compare two different sets of Bartik weights: Our baseline measure uses five year lags to compute the industry weights. For comparison, we keep the weights constant at the initial (1974) value. Both give the same qualitative answer, although the 1974 weights imply noisier responses that are smaller for new firms.

Figure F.4: Impulse-responses in baseline VAR: Baseline (5-year lagged) weights vs. constant 1974 weights.

(a) Startup shock



(b) Overall labor demand shock

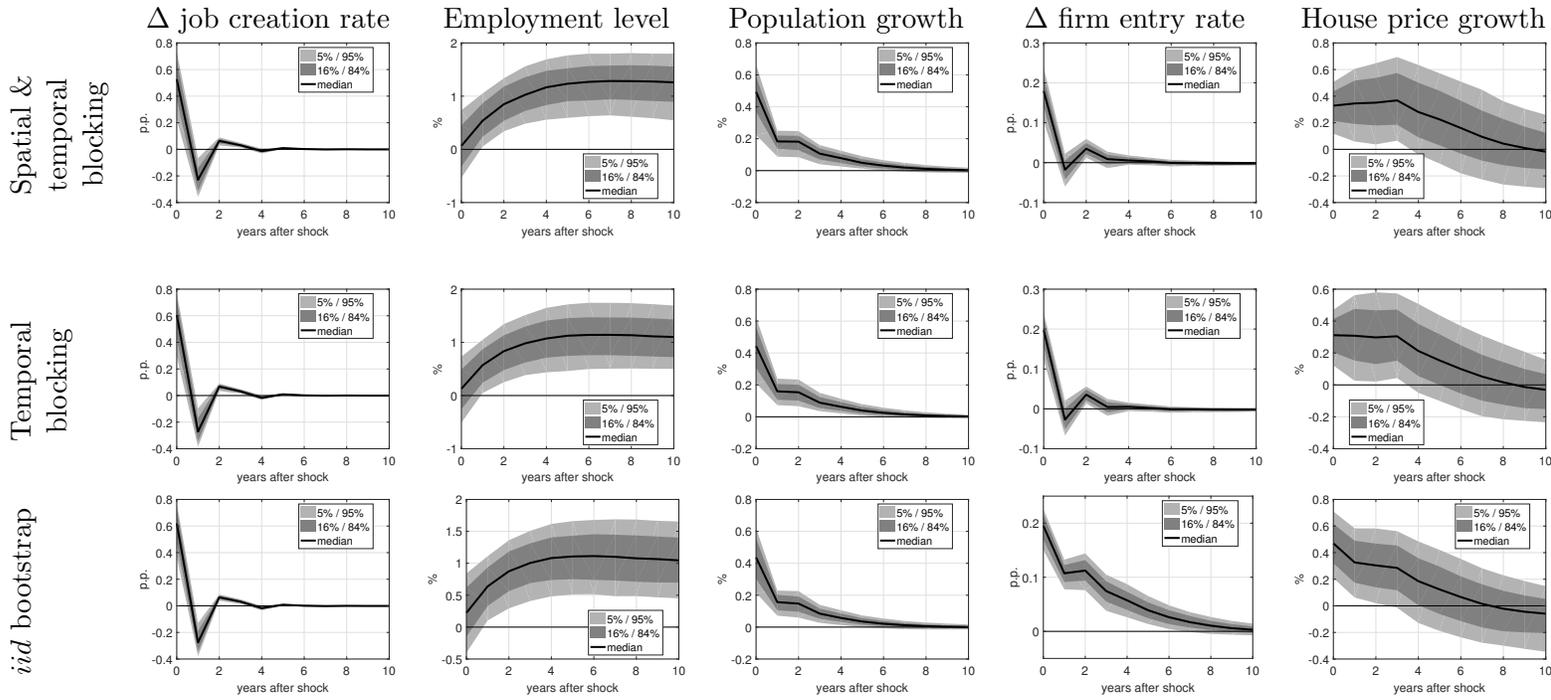


— CI median ■ 68% CI ■ 90% CI

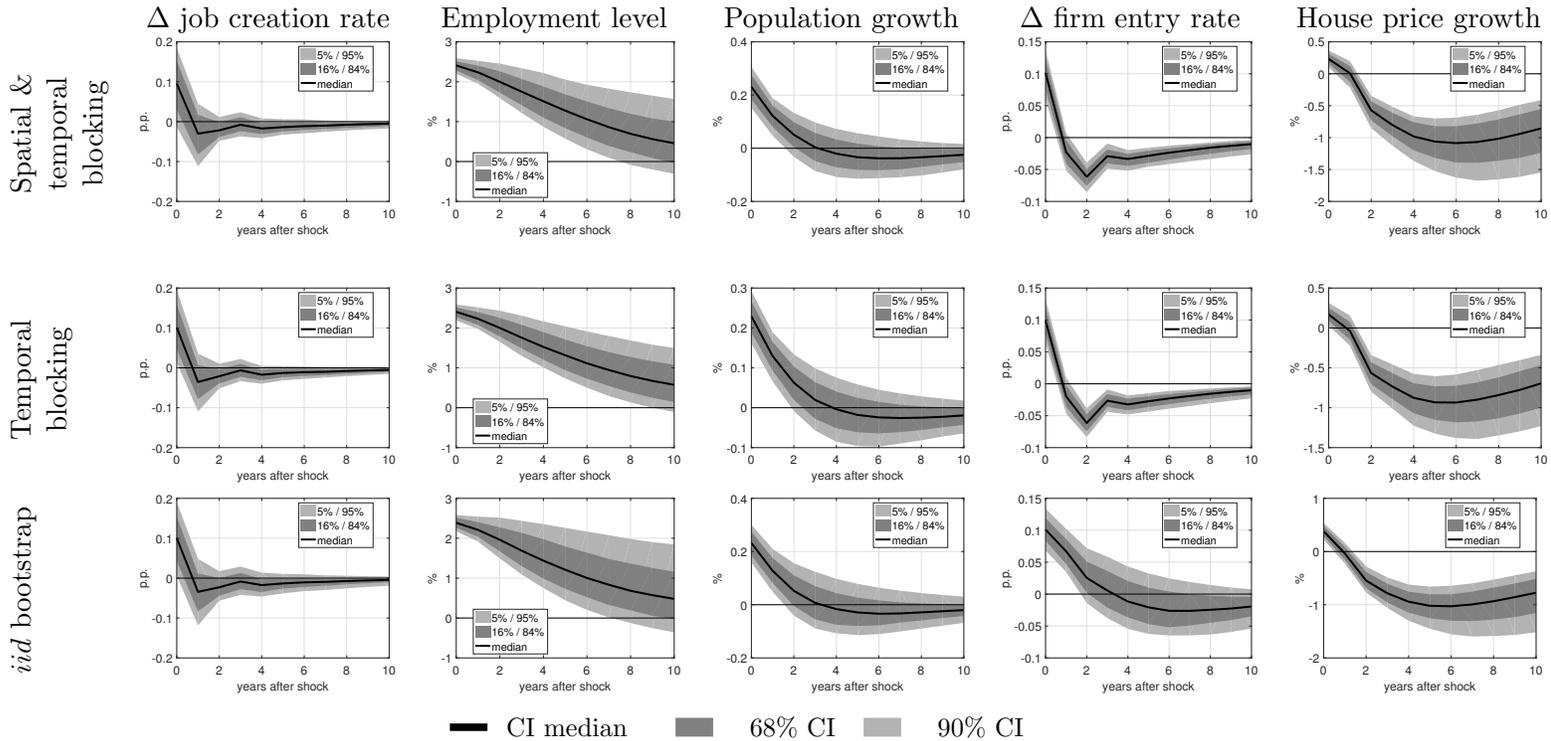
Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. The underlying house price data are from CoreLogic Solutions. TFP growth is calculated using Computstat data. Copyright 2018, S&P Global Market Intelligence (and its affiliates, as applicable). Obtained via Wharton Research Data Services (WRDS). No further distribution and or reproduction permitted. Since we cannot subtract the own-MSA when calculating the Bartik instrument for our startup-Bartik and TFP-Bartik variable, we split the estimation into the 18 largest MSAs with more than 1% of the population at the beginning of our sample and the other 336 MSAs. Alternatively, we swap the employment-based Bartik instrument for the overall labor demand shock for an instrument that uses incumbents' TFP growth to instrument. Results change little.

Figure F.5: Impulse-responses in baseline VAR: Baseline vs dropping the largest MSAs with baseline instrument and with the TFP growth instrument. ⁵

(a) Startup shock



(b) Overall labor demand shock

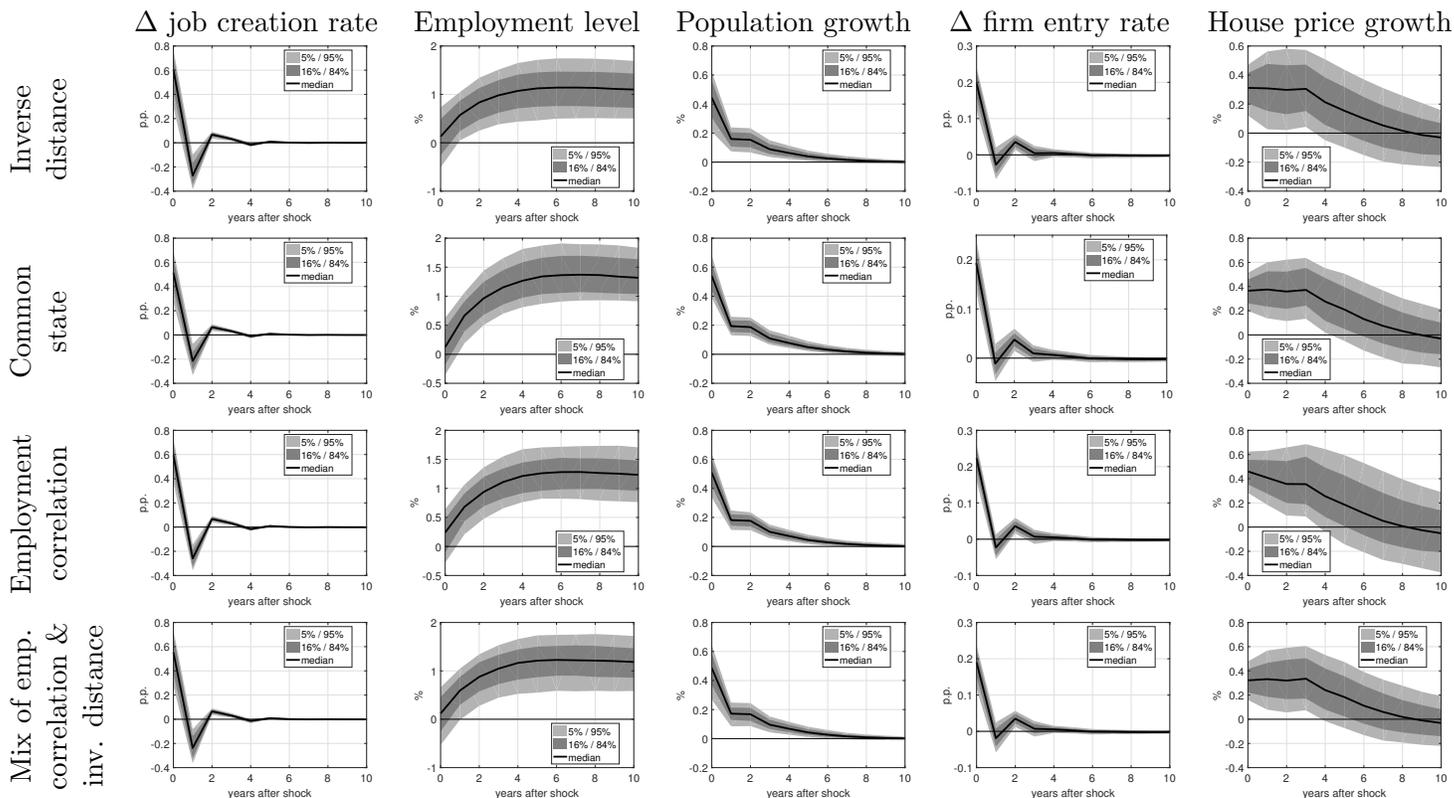


— CI median ■ 68% CI ■ 90% CI

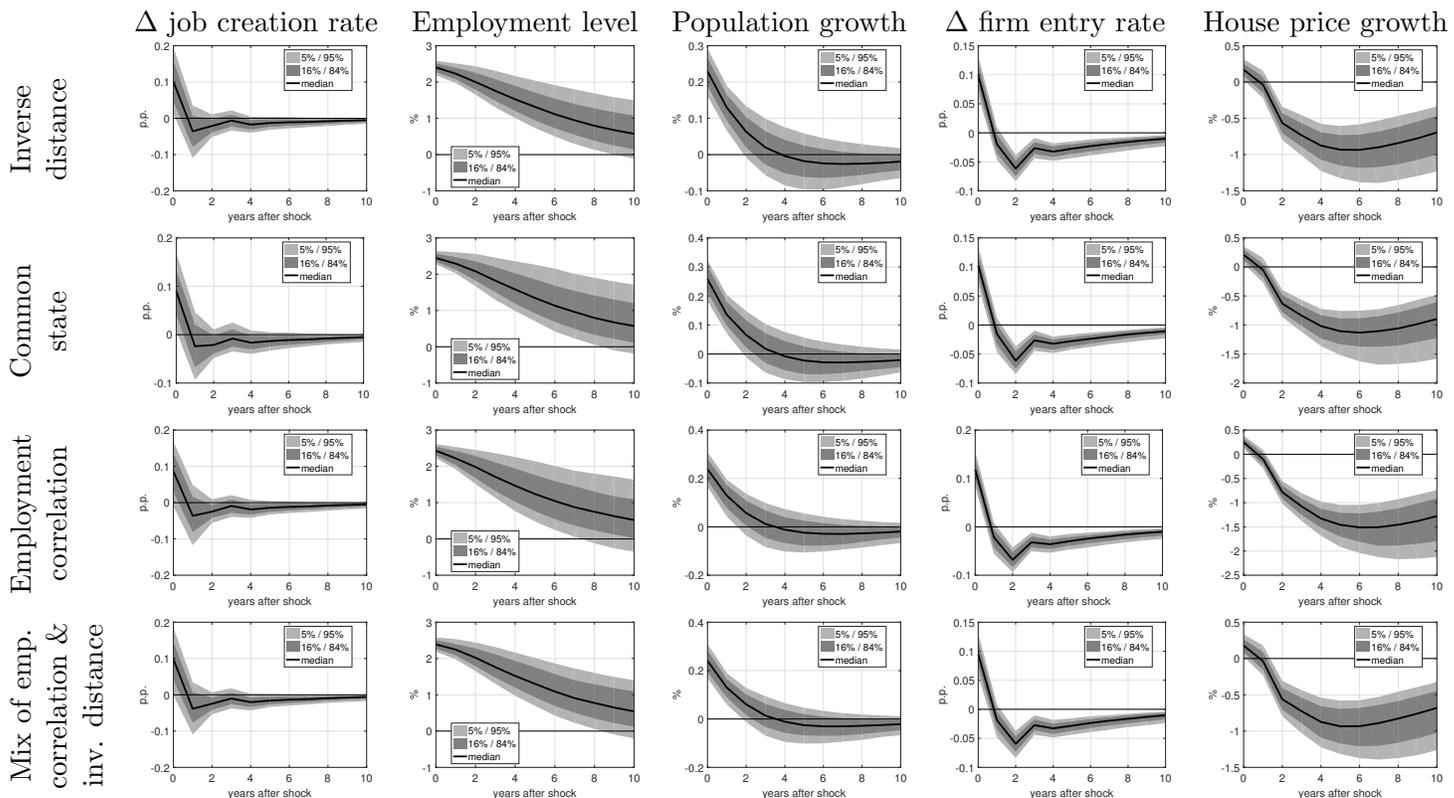
Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. The underlying house price data are from CoreLogic Solutions. We compare three bootstrap schemes: In our baseline, we sample blocks of three years at a time and impose the same shock realization on seven neighboring MSAs. In the *iid* bootstrap, we sample year by year independently. While the spatial correlation makes the results more uncertain, the effects are small.

Figure F.6: Impulse-responses in baseline VAR: 3-period block bootstrap (baseline) vs. *iid* bootstrap.

(a) Startup shock



(b) Overall labor demand shock



— CI median ■ 68% CI ■ 90% CI

See Figure 3 for a description of the plots. The underlying house price data are from CoreLogic Solutions. We compare proximity measures: The inverse Euclidian distance between MSA centroids (our baseline), defining neighbors using common states, and using the correlation of HP-filtered employment. We also estimate the best-fitting combination of the inverse distance and the employment correlation. All give similar answers.

Figure F.7: Impulse-responses in baseline VAR: Estimates based on different proximity measures.

Table F.3: Likelihood ratios for different proximity measures

$$2 \ln(L_c^E / L_c^S) / N$$

Specification	Inv. distance vs. common state	Inv. distance vs. correlation of employment	Inv. distance vs. estimated mix of state & inv. distance
Common spatial correlation	4.8	23.1	-4.1
Variable-specific spatial correlation	5.4	27.6	-3.8

The concentrated log-likelihood clearly favors the inverse Euclidean distance when considering a single proximity measure. With variable-specific spatial autocorrelation, the likelihood ratio is 5.4, and with common spatial correlation the ratio is 4.8. The differences are much larger when using the correlation of cyclical employment to compute correlations. The models are not nested, but the distance-based measure increases the fit significantly. The proximity measure that combines the inverse distance matrix and the common state matrix performs even better and assigns a 0.659 weight to the inverse distance matrix with a 90% confidence interval of (0.625, 0.695). House price data are from CoreLogic Solutions.

Table F.4: First-stage F -statistics in baseline VAR and large VAR

(a) Baseline

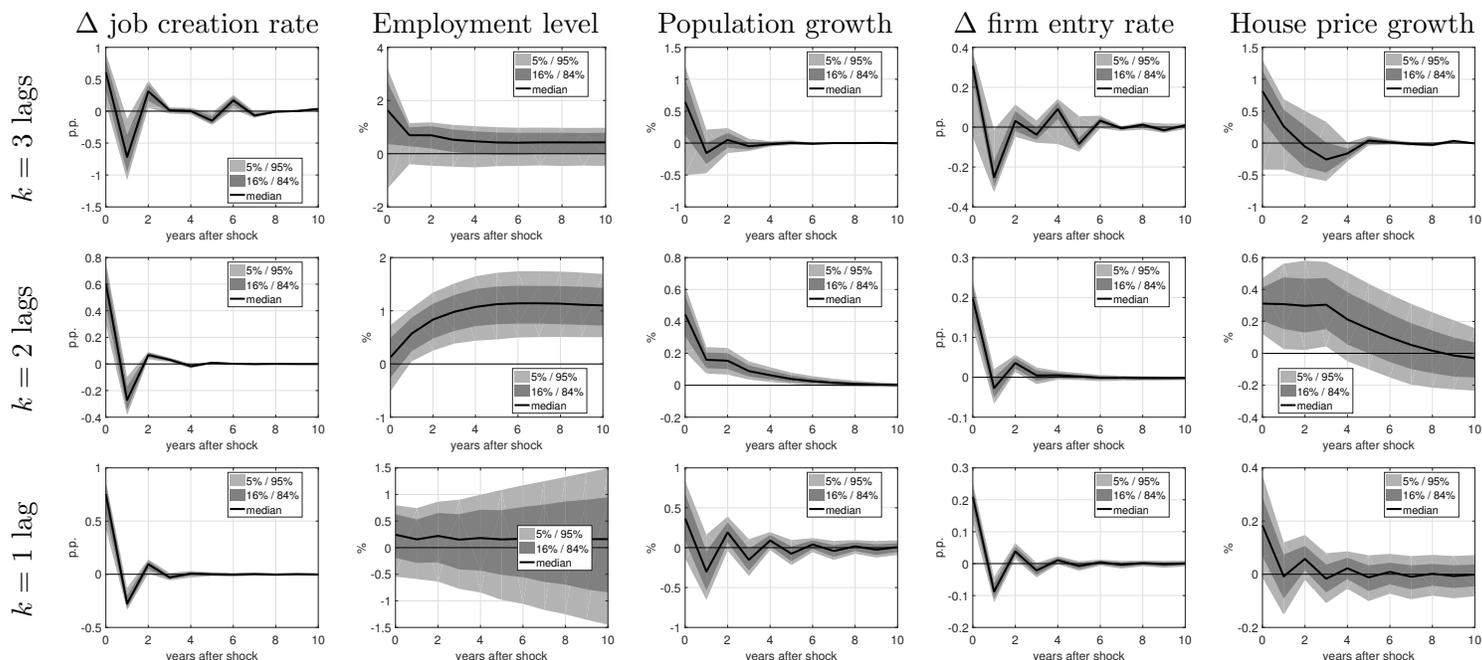
Variable	Point	Confidence interval				
	estimate	5%	16%	Median	84%	95%
Δ startup job creation rate	16.2	6.7	9.5	14.6	20.9	25.9
Employment/Pop ratio	84.3	32.5	48.0	72.3	93.6	106.1

(b) Large VAR

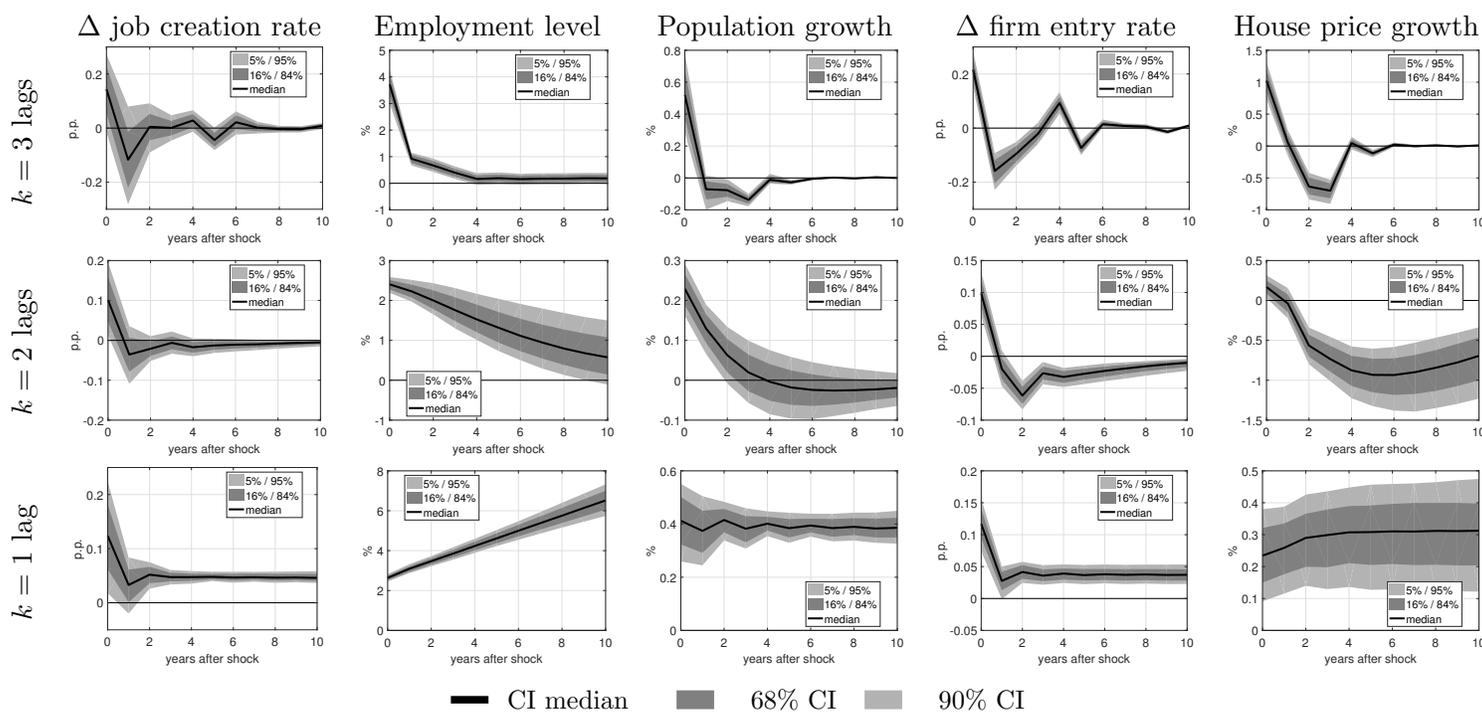
Variable	Point	Confidence interval				
	estimate	5%	16%	Median	84%	95%
Δ startup job creation rate	7.9	2.4	3.6	6.2	9.1	12.0
Employment/Pop ratio	43.6	12.4	20.5	36.6	47.7	55.2

The F -statistics that measure how well the instruments identify the structural shocks drop in the larger VAR. Intuitively, the identification problem becomes harder when we try to tell the two shocks apart from four other shocks, rather than two other shocks.

(a) Startup shock



(b) Overall labor demand shock

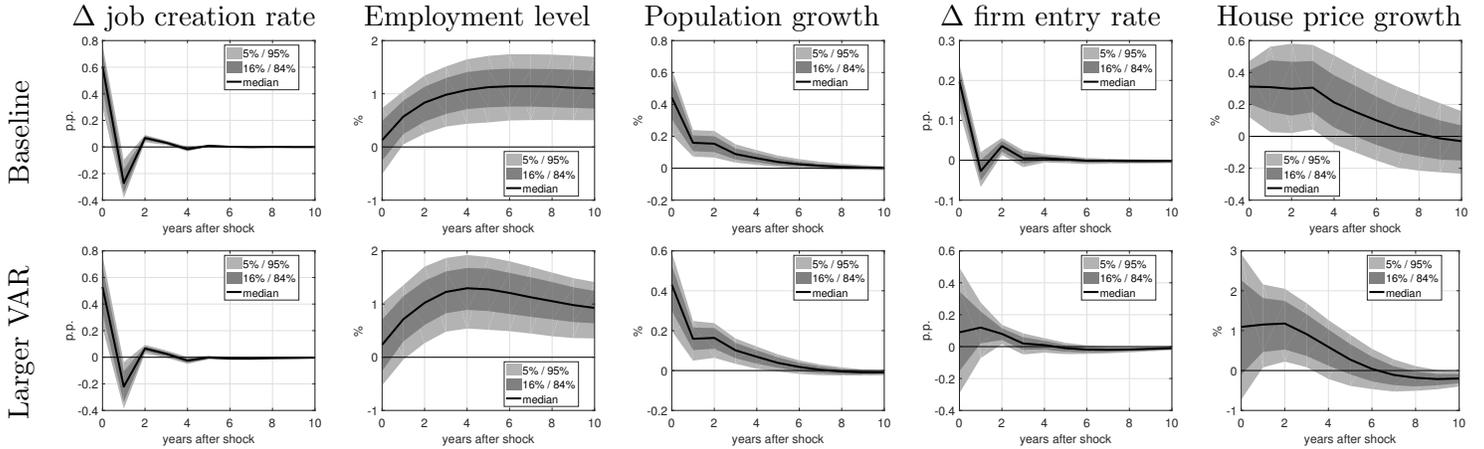


— CI median 68% CI 90% CI

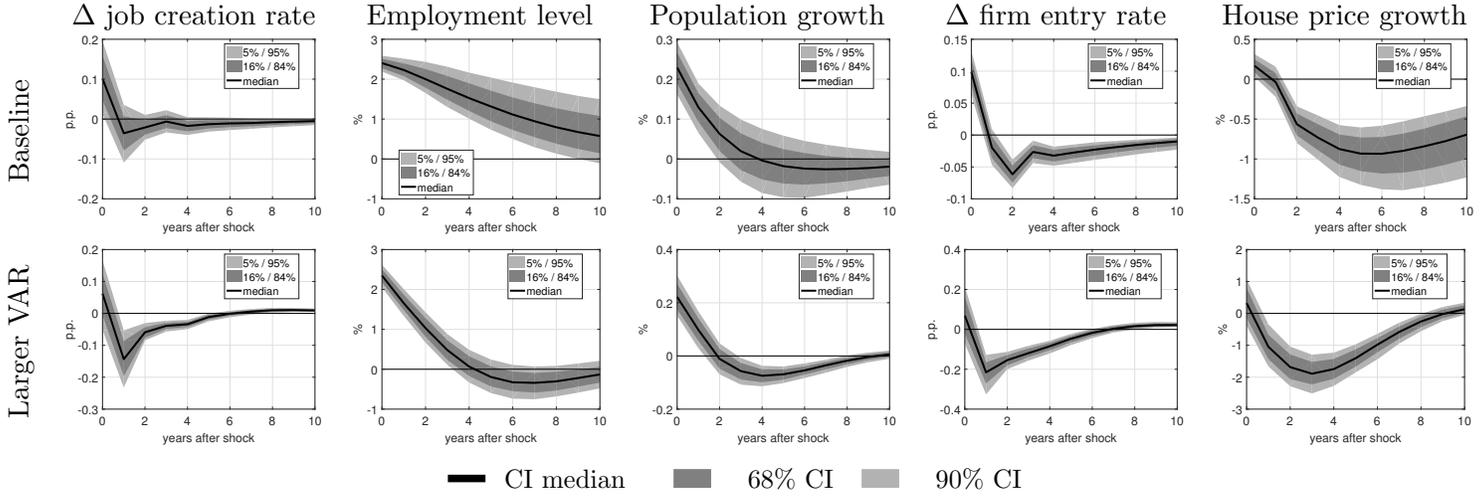
Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. The underlying house price data are from CoreLogic Solutions. We compare our baseline model with two lags to specifications with one or three lags. With a single lag, the VAR seems close to unstable and shock responses are qualitatively different from our baseline estimates. In contrast, with three lags we find qualitatively similar responses. Because theory suggests that we need a rich enough VAR specification, we conclude that we need at least two lags to capture the structural impulse-response functions well.

Figure F.8: Impulse-responses in baseline VAR: Comparing the lag length

(a) Startup shock

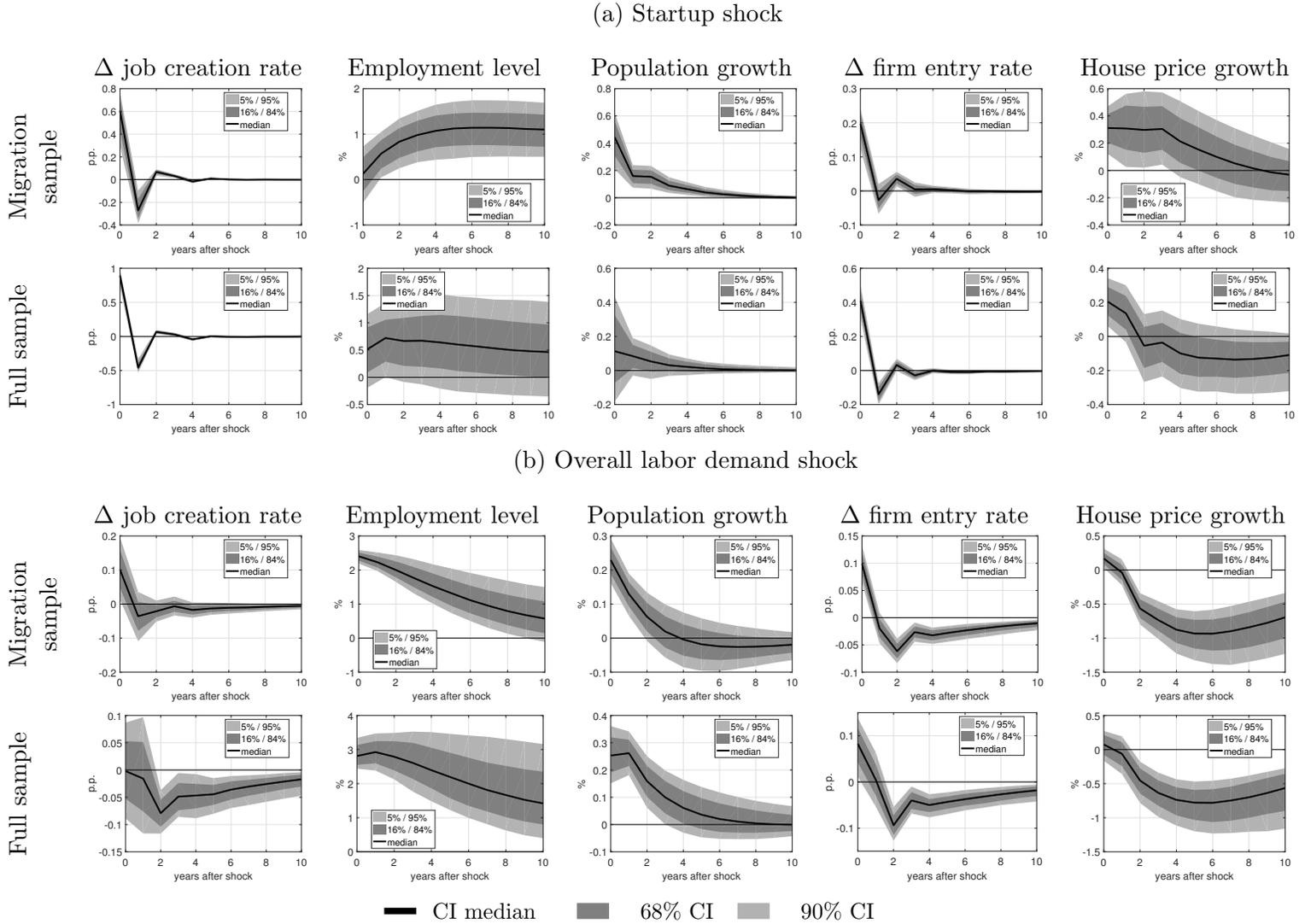


(b) Overall labor demand shock



Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. The underlying house price data are from CoreLogic Solutions. In our baseline VAR, we model firm entry rate, house price growth, and other variables without allowing them to feed back into the VAR. When we include entry and house prices in the core VAR, the responses of the original VAR variables change little. The responses of the entry rate and house price growth change slightly, but the larger model has less precise estimates whose confidence intervals are consistent with the estimates from the smaller model.

Figure F.9: Impulse-responses in baseline VAR and in larger VAR with entry rate and house prices.



Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. The underlying house price data are from CoreLogic Solutions. Here we drop the migration rate from the peripheral VAR to begin the estimation in 1980. We find qualitatively similar results, but noisier effects of the startup shock on employment and population growth.

Figure F.10: Impulse-responses in baseline VAR: Migration sample (1986–2013) vs. full sample (1981–2014).

Table F.5: Variance decomposition: Comparison of baseline and large wage VARs. Standard Bartik independent. 1986–2013. 2 lags. 68% Confidence interval

(a) Baseline VAR

	Startup shock		Overall labor demand		Other VAR shocks		Idiosyncratic shock	
Δ startup job creation rate	52.9	(27.6, 77.1)	2.0	(0.4, 3.7)	45.1	(21.6, 70.3)	0.0	(0.0, 0.0)
Employment/Pop ratio	3.6	(0.2, 7.0)	70.3	(63.5, 77.1)	26.1	(18.5, 33.4)	0.0	(0.0, 0.0)
Pop growth	39.8	(18.2, 62.2)	10.5	(6.6, 14.5)	49.6	(26.5, 70.7)	0.0	(0.0, 0.0)
Wage growth	6.7	(0.3, 12.9)	20.6	(14.7, 26.7)	72.7	(63.5, 82.5)	0.0	(0.0, 0.0)
Δ firm entry rate	6.8	(4.4, 9.1)	1.8	(1.1, 2.5)	2.5	(0.7, 4.3)	88.9	(87.4, 90.3)
House price growth	0.7	(0.3, 1.1)	0.2	(0.1, 0.4)	0.7	(0.3, 1.1)	98.4	(97.9, 98.9)
Δ young firm exit rate	0.1	(0.0, 0.2)	0.1	(0.0, 0.1)	0.1	(0.0, 0.2)	99.7	(99.6, 99.9)
Net migration rate	14.0	(6.2, 21.6)	6.2	(4.1, 8.0)	18.6	(9.7, 27.2)	61.2	(56.4, 65.9)
Firm exit rate (all)	0.3	(0.0, 0.6)	0.2	(0.1, 0.4)	1.2	(0.7, 1.8)	98.2	(97.5, 98.9)
Startup size	38.1	(18.8, 56.1)	1.1	(0.1, 2.1)	35.2	(17.4, 53.8)	25.6	(24.4, 26.9)

(b) Large VAR

	Startup shock		Overall labor demand		Other VAR shocks		Idiosyncratic shock	
Δ startup job creation rate	37.0	(14.5, 59.3)	1.0	(0.1, 2.0)	62.0	(39.6, 84.0)	0.0	(0.0, 0.0)
Employment/Pop ratio	3.2	(0.1, 6.7)	54.8	(45.6, 62.9)	42.1	(32.7, 51.9)	0.0	(0.0, 0.0)
Pop growth	36.9	(17.8, 55.3)	10.3	(5.8, 14.6)	52.8	(34.5, 72.0)	0.0	(0.0, 0.0)
Wage growth	5.1	(0.2, 10.3)	23.8	(16.7, 30.6)	71.1	(62.0, 80.1)	0.0	(0.0, 0.0)
Δ firm entry rate	6.2	(0.2, 11.8)	1.1	(0.0, 2.4)	92.7	(86.8, 99.0)	0.0	(0.0, 0.0)
House price growth	8.7	(0.5, 17.5)	1.1	(0.1, 2.3)	90.2	(81.5, 98.4)	0.0	(0.0, 0.0)
Firm exit rate (all)	0.8	(0.1, 1.6)	0.8	(0.3, 1.2)	3.7	(2.7, 4.8)	94.7	(93.7, 95.8)
Δ young firm exit rate	0.1	(0.0, 0.2)	0.1	(0.0, 0.2)	1.1	(0.7, 1.6)	98.6	(98.2, 99.1)
Net migration rate	14.7	(6.7, 22.4)	6.8	(4.2, 9.2)	23.1	(14.6, 31.5)	55.5	(51.3, 59.9)

Estimating a larger VAR leads to a similar variance decomposition for the variables that we model only in the periphery in our baseline but include in the larger VAR. The identified shocks still explain less than 15% of house price growth and less than 10% of firm entry. House price data are from CoreLogic Solutions.

Table F.6: First-stage F -statistics: Baseline VAR with various lag lengths

$k = 3$						
Variable	Point estimate	Confidence interval				
		5%	16%	Median	84%	95%
Δ startup job creation rate	18.4	2.0	3.0	5.1	7.4	9.1
Employment/Pop ratio	39.7	17.8	23.2	31.8	41.3	47.6
$k = 2$						
Variable	Point estimate	Confidence interval				
		5%	16%	Median	84%	95%
Δ startup job creation rate	16.2	6.7	9.5	14.6	20.9	25.9
Employment/Pop ratio	84.3	32.5	48.0	72.3	93.6	106.1
$k = 1$						
Variable	Point estimate	Confidence interval				
		5%	16%	Median	84%	95%
Δ startup job creation rate	14.0	4.1	6.7	11.3	16.6	20.3
Employment/Pop ratio	83.7	25.7	42.6	68.0	89.0	104.3

The F -statistics measuring the strength of the identification vary little with the number of lags included in the VAR and are always above 10.0. However, with three lags the bootstrapped distribution of the F -statistic shifts to the left.

F.3 Differences in initial density

Table F.7: Spatial autocorrelation: Coefficients estimates by variable; split by density

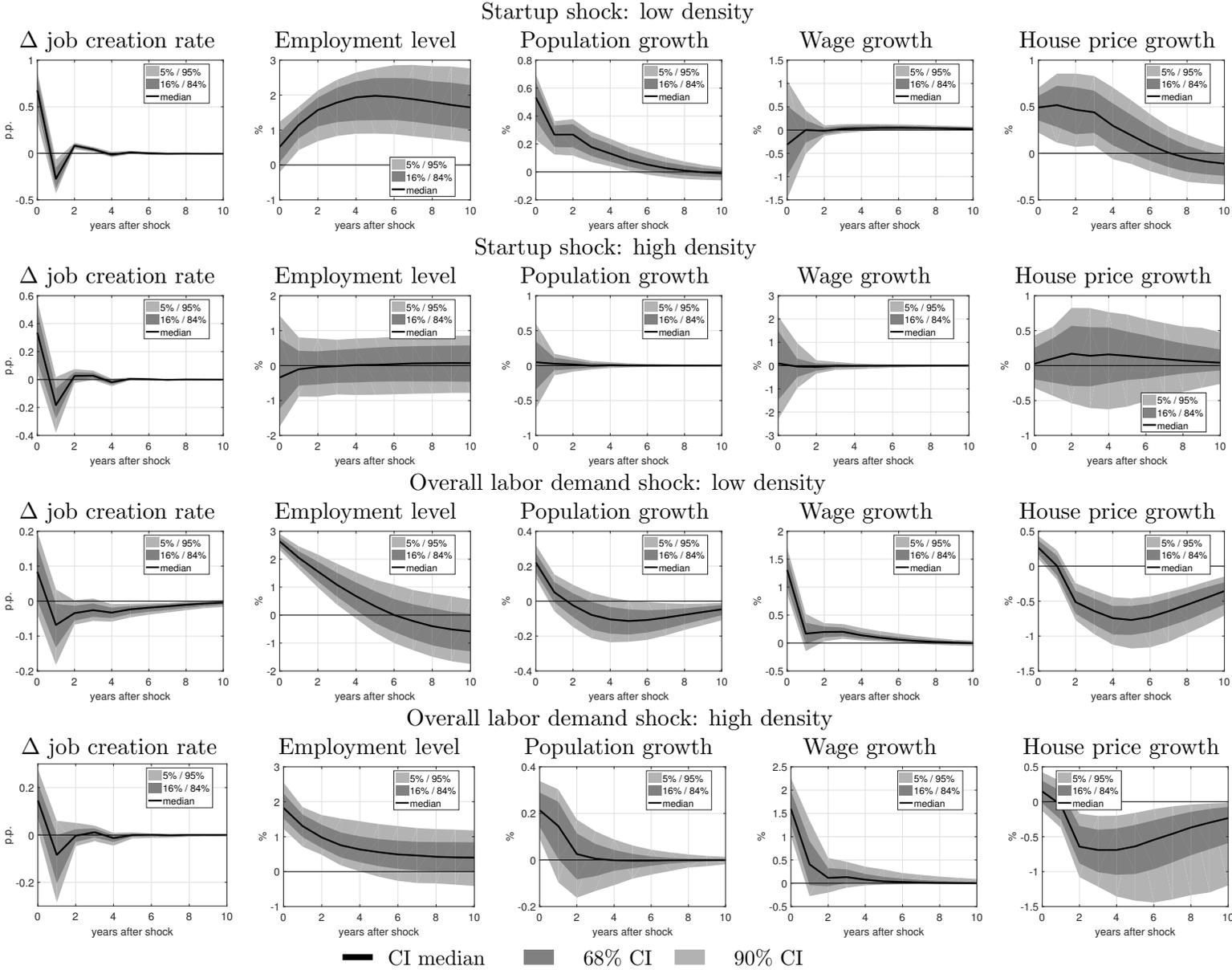
(a) Point estimates and confidence intervals for model with varying ρ s: Low density

Variable	Point	Avg.	Bias-corrected confidence interval				
	estimate	bias	5%	16%	Median	84%	95%
Gross job creation (births)	0.08	-0.02	0.01	0.03	0.06	0.09	0.12
Employment/Pop (log, BDS)	0.44	-0.05	0.34	0.36	0.39	0.42	0.44
Pop growth	0.52	-0.04	0.43	0.44	0.47	0.50	0.52
Wage growth	0.21	-0.02	0.14	0.16	0.19	0.22	0.24
Firm entry rate	0.50	-0.05	0.41	0.42	0.45	0.47	0.49
House price growth	0.81	-0.06	0.72	0.74	0.75	0.77	0.79
Firm exit rate	0.35	-0.05	0.26	0.28	0.30	0.33	0.34
Net migration rate	0.85	-0.09	0.70	0.73	0.76	0.78	0.80
Firm exit rate (all)	0.55	-0.05	0.46	0.47	0.50	0.52	0.54
Startup size	0.23	-0.04	0.15	0.16	0.19	0.21	0.23
Bartik: Entrant's job creation	0.85	-0.08	0.73	0.75	0.77	0.79	0.80
Bartik: Employment	0.57	-0.05	0.48	0.50	0.52	0.55	0.57

(b) Point estimates and confidence intervals for model with varying ρ s: High density

Variable	Point	Avg.	Bias-corrected confidence interval				
	estimate	bias	5%	16%	Median	84%	95%
Gross job creation (births)	0.15	-0.07	-0.00	0.03	0.08	0.14	0.20
Employment/Pop (log, BDS)	0.37	-0.05	0.18	0.24	0.31	0.37	0.41
Pop growth	0.71	-0.17	0.37	0.43	0.51	0.58	0.65
Wage growth	0.35	-0.11	0.14	0.19	0.24	0.30	0.33
Firm entry rate	0.45	-0.06	0.33	0.35	0.39	0.43	0.45
House price growth	0.81	-0.13	0.61	0.63	0.67	0.71	0.74
Firm exit rate	0.32	-0.07	0.17	0.20	0.25	0.30	0.32
Net migration rate	0.56	0.01	0.45	0.50	0.54	0.60	0.63
Firm exit rate (all)	0.63	-0.14	0.43	0.45	0.49	0.53	0.55
Startup size	0.38	-0.14	0.16	0.19	0.23	0.28	0.30
Bartik: Entrant's job creation	0.30	0.11	0.32	0.36	0.41	0.45	0.48
Bartik: Employment	0.57	-0.09	0.39	0.42	0.47	0.51	0.54

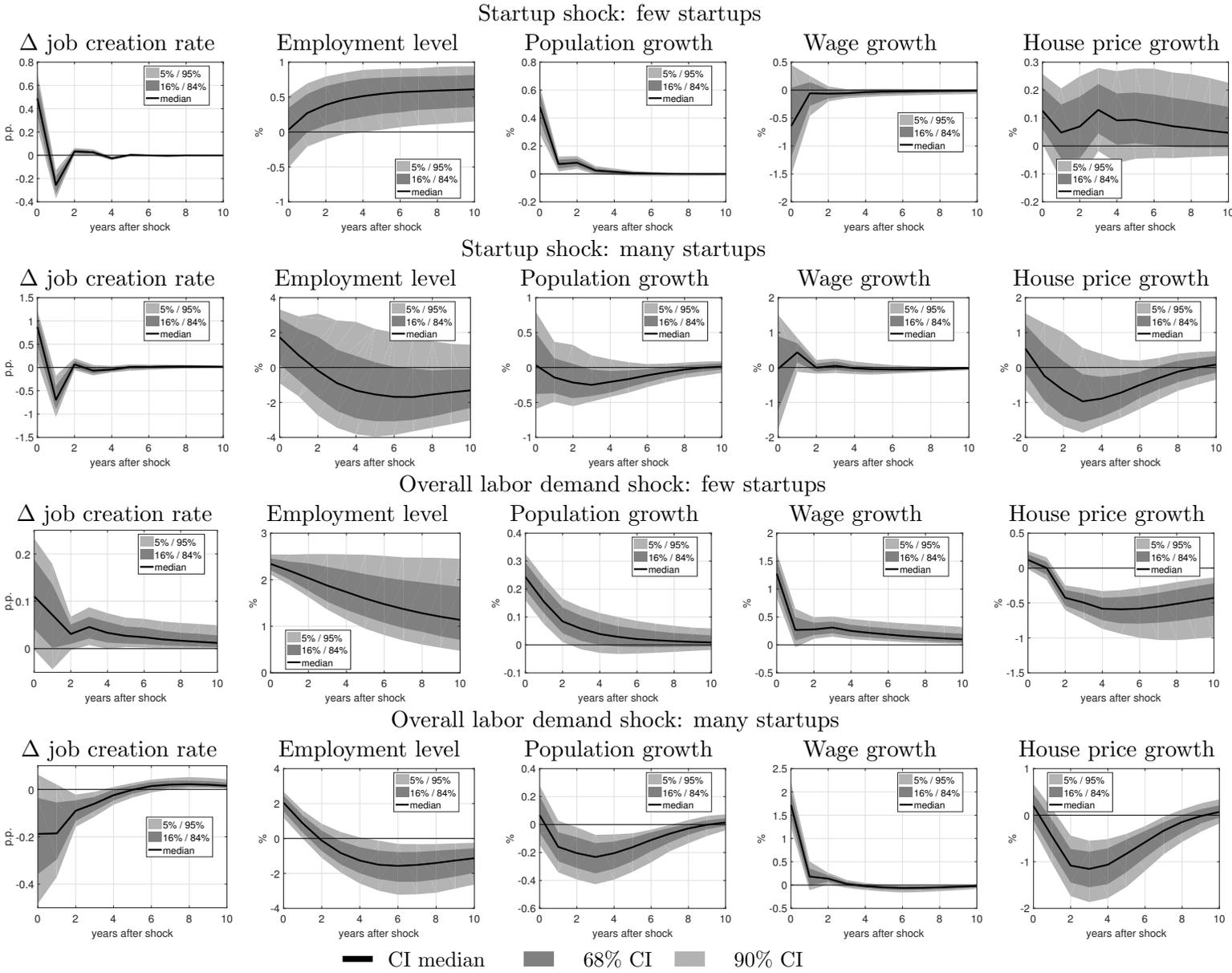
The estimated spatial correlation is small for low density cities, but larger than in the baseline estimates for high density areas. For high density MSAs, the spatial correlation varies significantly across variables. House price data are from CoreLogic Solutions.



Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. The underlying house price data are from CoreLogic Solutions. We split MSAs by their initial firm entry rate. The distribution is skewed to the right, and we set the cutoff at the 75th percentile. The effects of both shocks differ across MSAs with stronger effects in the MSAs with an initially lower startup rate. For MSAs with low initial startup rates, our results mirror our baseline estimates. For MSAs with many startups, our estimates of the effects of startup shocks are very noisy, and we find in Table F.9 that the corresponding F -statistic is low. We conclude that our identification is likely driven by MSAs with lower initial entry rates.

Figure F.11: Impulse-responses to startup and overall labor demand shocks for MSAs grouped by their initial density.

F.4 Differences in initial entry rate



Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. The underlying house price data are from CoreLogic Solutions. We split MSAs by their initial firm entry rate. The distribution is skewed to the right, and we set the cutoff at the 75th percentile. The effects of both shocks differ across MSAs with stronger effects in the MSAs with an initially lower startup rate. For MSAs with low initial startup rates, our results mirror our baseline estimates. For MSAs with many startups, our estimates of the effects of startup shocks are very noisy, and we find in Table F.9 that the corresponding F -statistic is low. We conclude that our identification is likely driven by MSAs with lower initial entry rates.

Figure F.12: Impulse-responses to startup and overall labor demand shocks for MSAs grouped by their initial entry rate.

Table F.8: Spatial autocorrelation: Coefficients estimates by variable; split by startup entry rate

(a) Point estimates and confidence intervals for model with varying ρ s: Low density

Variable	Point	Avg.	Bias-corrected confidence interval				
	estimate	bias	5%	16%	Median	84%	95%
Gross job creation (births)	0.05	0.05	0.03	0.05	0.09	0.16	0.23
Employment/Pop (log, BDS)	0.47	-0.07	0.36	0.38	0.42	0.46	0.51
Pop growth	0.58	-0.09	0.42	0.45	0.48	0.52	0.55
Wage growth	0.25	-0.01	0.18	0.21	0.26	0.33	0.39
Firm entry rate	0.45	-0.06	0.36	0.38	0.40	0.43	0.44
House price growth	0.85	-0.11	0.71	0.72	0.75	0.77	0.78
Firm exit rate	0.29	-0.04	0.21	0.22	0.25	0.28	0.30
Net migration rate	0.57	-0.04	0.49	0.50	0.54	0.56	0.58
Firm exit rate (all)	0.62	-0.08	0.49	0.51	0.53	0.56	0.57
Startup size	0.19	-0.04	0.14	0.17	0.22	0.27	0.31
Bartik: Entrant's job creation	0.75	-0.09	0.62	0.64	0.67	0.69	0.71
Bartik: Employment	0.57	-0.07	0.46	0.47	0.50	0.53	0.54

(b) Point estimates and confidence intervals for model with varying ρ s: High density

Variable	Point	Avg.	Bias-corrected confidence interval				
	estimate	bias	5%	16%	Median	84%	95%
Gross job creation (births)	0.09	-0.07	-0.06	-0.03	0.01	0.05	0.08
Employment/Pop (log, BDS)	0.19	-0.08	0.01	0.05	0.10	0.14	0.17
Pop growth	0.32	-0.09	0.16	0.19	0.23	0.27	0.30
Wage growth	0.12	-0.07	-0.03	0.00	0.05	0.09	0.13
Firm entry rate	0.49	-0.10	0.33	0.35	0.39	0.42	0.45
House price growth	0.66	-0.10	0.49	0.51	0.55	0.59	0.62
Firm exit rate	0.32	-0.11	0.14	0.17	0.21	0.25	0.28
Net migration rate	0.81	-0.14	0.58	0.61	0.66	0.70	0.73
Firm exit rate (all)	0.34	-0.04	0.24	0.26	0.30	0.33	0.36
Startup size	0.20	-0.07	0.05	0.08	0.12	0.15	0.17
Bartik: Entrant's job creation	0.46	-0.07	0.32	0.35	0.39	0.43	0.46
Bartik: Employment	0.38	-0.07	0.23	0.26	0.31	0.35	0.38

The estimated spatial correlation is small for MSAs with few startups, but comparable to the baseline estimates for areas with high entry rates. For high entry MSAs, the spatial correlation varies significantly across variables. House price data are from CoreLogic Solutions.

Table F.9: First-stage F -statistics split by initial entry rates

(a) Low initial entry rate

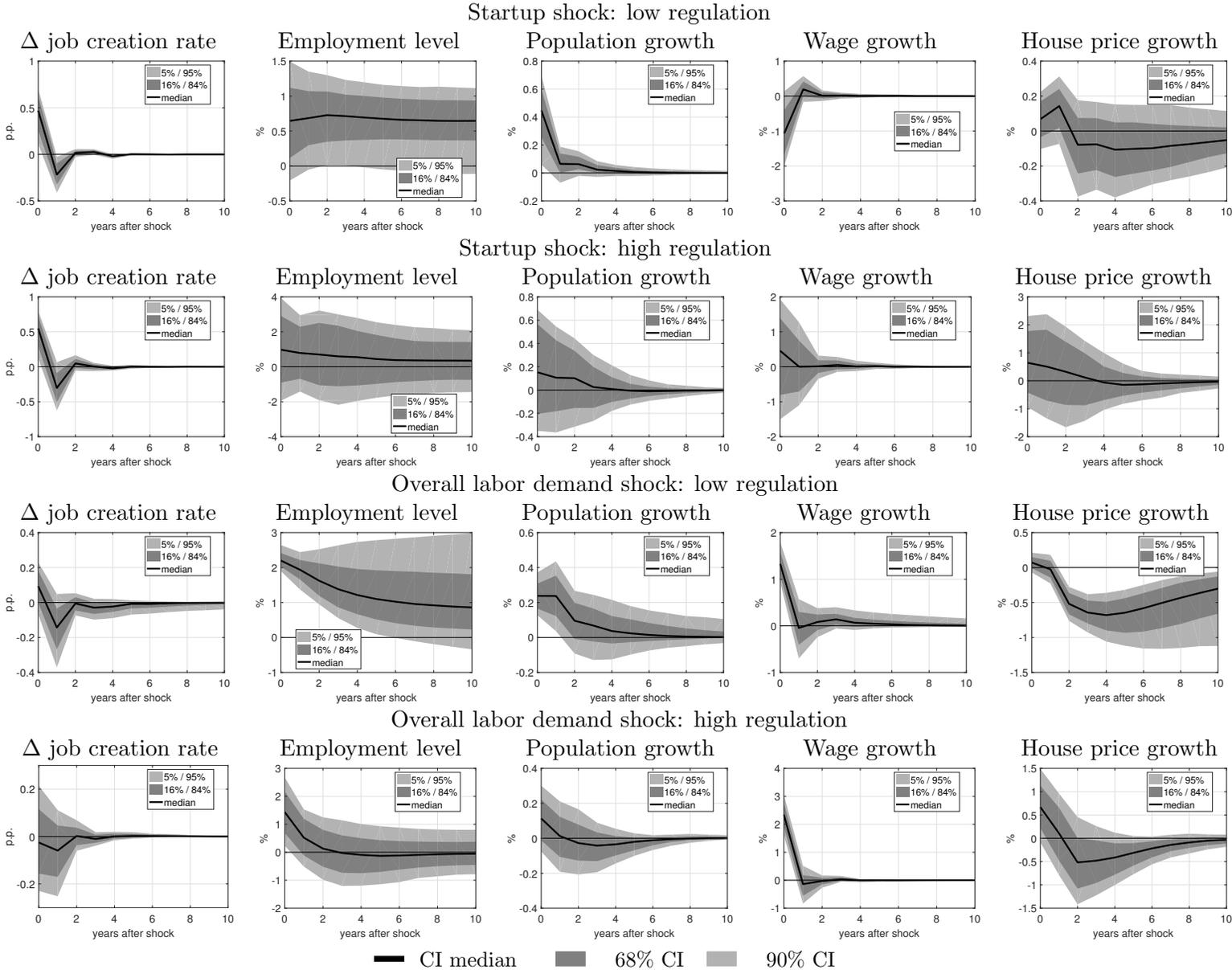
	Point	Confidence interval				
Δ startup job creation rate	10.2	5.8	8.9	14.9	21.4	26.9
Employment/Pop ratio	29.5	20.6	31.4	45.1	61.6	74.4

(b) High initial entry rate

	Point	Confidence interval				
Δ startup job creation rate	6.1	2.1	3.5	6.5	10.3	13.5
Employment/Pop ratio	26.9	13.5	17.9	24.6	36.0	42.1

The F -statistics measuring the strength of the identification indicate that our instruments predict the identified shocks well, except for startup shocks in high entry MSAs.

F.5 Differences in Wharton Regulation Index



Δ job creation rate refers to the job creation rate by startups. Panel (a) shows the response to the identified startup shock, along with bootstrapped confidence intervals. Panel (b) shows the corresponding response to the overall labor demand shock. The solid line is the median across the bootstrapped draws, while the shaded areas are the 68% and 90% confidence intervals, respectively. The underlying house price data are from CoreLogic Solutions. We split MSAs by their initial firm entry rate. The distribution is skewed to the right, and we set the cutoff at the 67th percentile. The effects of both shocks differ across MSAs with stronger effects in the MSAs with an initially lower startup rate. For MSAs with low regulation, our results are similar to our baseline estimates, but stronger. For MSAs with high regulation, our estimates of the effects of startup shocks are very noisy, and we find in Table F.11 that the corresponding F -statistic is low. We conclude that our identification is likely driven by MSAs with lower regulation. Higher regulation MSAs, however, also show a weaker response to overall labor demand shocks.

Figure F.13: Impulse-responses to startup and overall labor demand shocks for MSAs grouped by their Wharton Regulation Index number.

Table F.10: Spatial autocorrelation: Coefficients estimates by variable; split by Wharton Regulation Index

(a) Point estimates and confidence intervals for model with varying ρ s: Low density

Variable	Point estimate	Avg. bias	Bias-corrected confidence interval				
			5%	16%	Median	84%	95%
Gross job creation (births)	0.11	-0.00	0.03	0.07	0.13	0.24	0.33
Employment/Pop (log, BDS)	0.38	-0.07	0.24	0.28	0.34	0.42	0.49
Pop growth	0.76	-0.18	0.50	0.53	0.58	0.64	0.70
Wage growth	0.19	-0.05	0.07	0.10	0.16	0.23	0.31
Firm entry rate	0.38	-0.06	0.27	0.29	0.33	0.37	0.39
House price growth	0.80	-0.13	0.61	0.64	0.66	0.70	0.72
Firm exit rate	0.33	-0.10	0.18	0.21	0.24	0.28	0.30
Net migration rate	0.86	-0.15	0.64	0.67	0.71	0.74	0.76
Firm exit rate (all)	0.44	-0.07	0.31	0.33	0.36	0.40	0.42
Startup size	0.29	-0.09	0.15	0.18	0.23	0.28	0.33
Bartik: Entrant's job creation	0.83	-0.14	0.64	0.66	0.70	0.73	0.75
Bartik: Employment	0.54	-0.10	0.38	0.40	0.44	0.49	0.51

(b) Point estimates and confidence intervals for model with varying ρ s: High density

Variable	Point estimate	Avg. bias	Bias-corrected confidence interval				
			5%	16%	Median	84%	95%
Gross job creation (births)	0.12	-0.07	-0.03	0.01	0.06	0.12	0.14
Employment/Pop (log, BDS)	0.35	-0.08	0.15	0.18	0.24	0.30	0.34
Pop growth	0.22	0.02	0.12	0.16	0.22	0.28	0.32
Wage growth	0.16	-0.07	-0.01	0.02	0.08	0.13	0.17
Firm entry rate	0.57	-0.13	0.37	0.40	0.43	0.47	0.49
House price growth	0.68	-0.09	0.54	0.56	0.59	0.63	0.65
Firm exit rate	0.35	-0.07	0.22	0.24	0.28	0.32	0.35
Net migration rate	0.76	-0.14	0.55	0.58	0.63	0.67	0.69
Firm exit rate (all)	0.61	-0.14	0.41	0.43	0.47	0.50	0.53
Startup size	0.29	-0.06	0.14	0.18	0.22	0.27	0.30
Bartik: Entrant's job creation	0.55	-0.05	0.45	0.48	0.51	0.55	0.57
Bartik: Employment	0.34	-0.04	0.24	0.26	0.31	0.35	0.38

The estimated spatial correlation is slightly negative for MSAs with low regulation, but comparable to the baseline estimates for areas with high regulation. For both high and low regulation MSAs, the spatial correlation varies significantly across variables. House price data are from CoreLogic Solutions.

Table F.11: First-stage F -statistics split by Wharton Regulation Index

(a) Low Wharton Regulation Index						
	Point	Confidence interval				
Δ startup job creation rate	11.2	3.7	5.5	10.0	15.5	21.1
Employment/Pop ratio	51.6	17.2	24.5	36.6	51.8	63.1
(b) High Wharton Regulation Index						
	Point	Confidence interval				
Δ startup job creation rate	5.0	1.5	2.4	4.8	7.5	10.3
Employment/Pop ratio	13.6	6.5	9.1	14.6	21.5	27.3

The F -statistics measuring the strength of the identification indicate that our instruments predict the identified shocks well, except for startup shocks in highly regulated MSAs.

Disclosures

Gerald Carlino has nothing to disclose.

Thorsten Drautzburg has nothing to disclose.