**Remote Work & Rural–Urban Differences in Labor Force Participation**

**Geography:** Iowa counties  
**Period:** 2021–2024 (annual averages)  
**Author:** Sam Queen (with ChatGPT collaboration)

**Table of Contents**

1. Research Question & Motivation
2. Conceptual Framework: Why Remote Work Might Affect LFPR
3. Units of Analysis & Time Structure
4. Core Variables & Notation
5. Data Sources Overview
6. ACS + PEP Hybrid Method for Dynamic County 16+ Populations
7. LFPR Construction Algorithm (County-Year)
8. Remote Work Data Integration (Survey-Based)
9. Rural / Urban Classification Options
10. Recommended Inference Techniques
11. Estimation Issues & Robustness
12. Modeling Sequence Recommendation
13. R Implementation Skeleton
14. Data Quality Diagnostics
15. Reporting & Visualization Ideas
16. Extensions & Sensitivity Analyses
17. Reproducibility & Documentation Checklist

**1. Research Question & Motivation**

**Primary question:** *How has remote work affected labor force participation differently in rural vs. urban areas in Iowa from 2021 through 2024?*

**Sub-questions**

* Did remote work adoption help retain or attract labor force participants in rural counties?
* Are urban counties seeing substitution (workers leaving the labor force, switching to remote self-employment)?
* Do broadband availability, industry mix, or education levels mediate the remote work → LFPR relationship?

**Why it matters**

* Post-pandemic labor market shifts are uneven across space.
* Iowa policymakers track rural vitality, workforce shortages, and out-migration.
* Remote-capable jobs could decouple residence from employment centers, altering participation capacity in rural labor markets.

**2. Conceptual Framework: Why Remote Work Might Affect LFPR**

Potential channels linking remote work to labor force participation (LFPR):

| **Channel** | **Direction** | **Rural vs. Urban Salience** | **Notes** |
| --- | --- | --- | --- |
| Commute cost reduction | ↑ LFPR | Stronger rural (longer commutes avoided) | Encourages marginal entrants (parents, older workers). |
| Childcare flexibility | ↑ LFPR | Mixed; depends on home support | Remote/hybrid increases ability to work part-time. |
| Health risk avoidance | ↑ LFPR | Pandemic waves; older rural workers | Work-from-home lowers exposure. |
| Broadband constraint | ↓ Potential | Strong rural | Remote work impossible without connectivity. |
| Amenity migration | Ambiguous | Urban → rural moves | Could raise rural pop16+ and LFPR denominator dynamics. |
| Industry composition | Varies | Urban knowledge jobs more remote-capable | Structural differences matter; include controls. |

This framework motivates interaction terms (Remote × Rural) and mediation/pathway analyses.

**3. Units of Analysis & Time Structure**

* **Spatial unit:** Iowa *county* (all 99).
* **Temporal unit:** Calendar year (annual averages).
* Underlying data (LAUS monthly, survey waves, etc.) can be aggregated to annual to stabilize variance.
* Balanced panel expected: 99 counties × 4 years = 396 observations (before missingness).

**4. Core Variables & Notation**

Let county index cc, year index t∈{2021,2022,2023,2024}t\in\{2021,2022,2023,2024\}.

**Labor market quantities**

* LFctLF\_{ct}: Labor force (Employed + Unemployed) from BLS LAUS annual averages.
* EctE\_{ct}, UctU\_{ct} if needed.

**Population**

* POP16c,2021ACSPOP16^{ACS}\_{c,2021}: Baseline ACS 5-year estimate of civilian non-institutional population age 16+ in 2021.
* POPc,tPEPPOP^{PEP}\_{c,t}: Total population (all ages) from Census PEP for each year.
* POP16ctPOP16\_{ct}: Estimated 16+ population in year tt via ACS+PEP hybrid (see Section 6).

**Remote work**

* RWctRW\_{ct}: Share of workers in county-year working remotely (from your survey; see Section 8).

**Rurality**

* RURcRUR\_c: Indicator (1=rural, 0=urban) or multi-level categorical (e.g., RUCC groups).

**Labor Force Participation Rate**

LFPRct=100×LFctPOP16ct.LFPR\_{ct} = 100 \times \frac{LF\_{ct}}{POP16\_{ct}}.

**5. Data Sources Overview**

Below are the data elements you will merge to build the analytic panel.

**5.1 Core for LFPR Construction**

**BLS Local Area Unemployment Statistics (LAUS)**

* Annual county-level labor force, employment, unemployment.
* Download annual averages; or pull monthly and average.

**ACS 5-year Estimates (baseline 2021)**

* Table B23001 (employment status by sex & age) or B23025 (employment status for population 16+).
* Use for POP16c,2021ACSPOP16^{ACS}\_{c,2021}.

**Census Population Estimates Program (PEP)**

* Annual July 1 county total population estimates, 2021–2024.
* Use for growth scaling.

**5.2 Remote Work Exposure**

**Your survey data** (has respondent county & remote classification).

* Aggregate to RWctRW\_{ct} with weights if available.

**5.3 Rural / Urban Classification**

Options (see Section 9):

* USDA ERS RUCC (codes 1–9).
* RUCA codes (Commuting-based).
* OMB Metropolitan / Micropolitan definitions.

**5.4 Control Variables (Optional but Recommended)**

| **Construct** | **Candidate Source** | **Notes** |
| --- | --- | --- |
| Broadband access % households | FCC Broadband Data Collection; NTIA | Lag & quality issues—smooth over years. |
| Industry mix (shares emp.) | QCEW, ACS, BEA REIS | Extract remote-capable industry share. |
| Educational attainment | ACS (B15003) | % BA+ often correlated with remote adoption. |
| Age structure | ACS | Support for older worker LFPR. |
| Median household income | ACS | Local demand conditions proxy. |

**6. ACS + PEP Hybrid Method for Dynamic County 16+ Populations**

**Problem:** LAUS gives labor force counts, but not the denominator (civilian 16+) at high frequency. ACS is lagged; PEP has totals but no 16+ breakout.  
**Solution:** Anchor 16+ in 2021 (ACS), scale forward using PEP total population growth.

**6.1 Notation**

* POPc,tPEPPOP^{PEP}\_{c,t}: Total population (all ages) in county cc, year tt.
* Baseline year: 2021.
* Horizon: 2021–2024.

**6.2 Compute County-Specific Compound Annual Growth Rate (CAGR)**

gc=(POPc,2024PEPPOPc,2021PEP)1/3−1.g\_c = \left(\frac{POP^{PEP}\_{c,2024}}{POP^{PEP}\_{c,2021}}\right)^{1/3} - 1.

Rationale: 3 intervals (2021→2022, 2022→2023, 2023→2024). Assumes smooth compounding.

**6.3 Generate Year-Specific Growth Factors**

For any year tt:

GFc,t=(1+gc)(t−2021).GF\_{c,t} = (1 + g\_c)^{(t-2021)}.

By definition: GFc,2021=1GF\_{c,2021} = 1; GFc,2024=(1+gc)3GF\_{c,2024} = (1+g\_c)^3 ≈ POPc,2024PEP/POPc,2021PEPPOP^{PEP}\_{c,2024}/POP^{PEP}\_{c,2021}.

**6.4 Estimate 16+ Population by Year**

POP16c,t=POP16c,2021ACS×GFc,t.POP16\_{c,t} = POP16^{ACS}\_{c,2021} \times GF\_{c,t}.

Interpretation: The 16+ share of population is treated as stable over 3 years; total change moves proportionally with PEP population trends.

**6.5 Optional Refinements**

**Age drift adjustment:** If ACS 2022/2023 5-year releases available, update proportion 16+ using linear interpolation between ACS series.  
**Negative or Extreme Growth:** If POPc,2024PEP<POPc,2021PEPPOP^{PEP}\_{c,2024} < POP^{PEP}\_{c,2021} (population decline), CAGR will be negative; formulas still hold. Consider floor at zero for very small counties if numeric instabilities appear.  
**Alternative (Year-Specific Growth):** Instead of CAGR, use annual PEP growth ratios year-to-year:  
rc,2022=POPc,2022PEP/POPc,2021PEPr\_{c,2022} = POP^{PEP}\_{c,2022} / POP^{PEP}\_{c,2021}, etc., then scale baseline 16+ by each ratio—captures non-linear population paths but may import noise.

**7. LFPR Construction Algorithm (County-Year)**

**Inputs:** LAUS LFctLF\_{ct}; estimated POP16ctPOP16\_{ct} from Section 6.

**Algorithm:**

1. Import LAUS annual county labor force counts (E, U, LF).
2. Import ACS 2021 16+ population.
3. Import PEP 2021–2024 total population; compute growth factors.
4. Estimate POP16ctPOP16\_{ct}.
5. Compute LFPR: LFPRct=100∗LFct/POP16ctLFPR\_{ct} = 100 \* LF\_{ct} / POP16\_{ct}.
6. (Quality check) Flag counties where implied LFPR > 85% or < 35% (inspect denominators).

**8. Remote Work Data Integration (Survey-Based)**

You have micro survey responses with county and remote classification.

**8.1 Define Remote Work Flag**

* Binary: Remote if working primarily from home (or above threshold % remote days).
* Alternative: Multi-category (On-site / Hybrid / Fully Remote) collapsed to numeric scale (0–1).

**8.2 Weighting**

* If survey has person weights, use them.
* If not, post-stratify to ACS industry × education × rural group where sample allows.

**8.3 Aggregate to County-Year**

For each county-year:

RWct=∑iwiRemotei∑iwi,RW\_{ct} = \frac{\sum\_i w\_i Remote\_i}{\sum\_i w\_i},

where wiw\_i are survey weights (default 1 if none).

If sample too thin for small rural counties, pool across adjacent counties or use empirical Bayes shrinkage toward RUCC group means.

**9. Rural / Urban Classification Options**

**Pick one primary specification; test robustness with alternatives.**

| **Scheme** | **Data** | **Coding Idea** | **Pros** | **Cons** |
| --- | --- | --- | --- | --- |
| USDA RUCC (1–9) | County | Collapse 1–3 Urban, 4–9 Rural | Widely used in ag/rural research | Coarse; commuting patterns not captured well. |
| RUCA | ZIP / Tract | Map to county shares | Captures commuting urbanization | Mapping step required. |
| OMB Metro/Micro | CBSA | Binary metro indicator | Simple | Misses gradations. |

**Recommendation:** Use RUCC collapsed to binary for main model; include multi-level category in robustness checks.

**10. Recommended Inference Techniques**

Below are complementary strategies; you can mix descriptive and regression-based approaches.

**10.1 Descriptive Grouped Means**

* Compare mean LFPR Rural vs. Urban each year.
* Welch t-tests (unequal variances).
* Test change over time: compare ΔLFPR2021→2024\Delta LFPR\_{2021\to2024} rural vs. urban.

**10.2 Two-Way Fixed Effects Panel Regression (Core Model)**

LFPRct=α+β1RWct+β2(RWct×RURc)+γ′Xct+μc+λt+ϵct.LFPR\_{ct} = \alpha + \beta\_1 RW\_{ct} + \beta\_2 (RW\_{ct} \times RUR\_c) + \gamma' X\_{ct} + \mu\_c + \lambda\_t + \epsilon\_{ct}.

* μc\mu\_c: County fixed effects (absorbs time-invariant rurality differences; you can omit main RUR if FE used).
* λt\lambda\_t: Year fixed effects (statewide shocks).
* Interpretation: β1\beta\_1 is remote effect in urban counties (baseline); β2\beta\_2 is *additional* effect in rural counties.
* Cluster SEs at county level.

**10.3 Difference-in-Differences (High-Remote vs. Low-Remote Counties)**

Define treated counties as those above median remote share increase from 2021 baseline.  
Model:

LFPRct=α+δTreatc+θPostt+τ(Treatc×Postt)+μc+λt+ϵct.LFPR\_{ct} = \alpha + \delta Treat\_c + \theta Post\_t + \tau (Treat\_c \times Post\_t) + \mu\_c + \lambda\_t + \epsilon\_{ct}.

* Choose Post period (e.g., 2023–2024) when remote stabilized.
* τ\tau estimates differential change in LFPR among high-remote counties.
* Add rural interactions: (Treatc×Postt×RURc)(Treat\_c \times Post\_t \times RUR\_c) for triple-diff.

**10.4 Event Study / Dynamic Remote Effects**

If you have annual or sub-annual RWctRW\_{ct}, estimate leads/lags:

LFPRct=α+∑k=−KLβkDc,t+kRemoteShock+μc+λt+Xct+ϵct.LFPR\_{ct} = \alpha + \sum\_{k=-K}^{L} \beta\_k D\_{c,t+k}^{RemoteShock} + \mu\_c + \lambda\_t + X\_{ct} + \epsilon\_{ct}.

Plot βk\beta\_k to inspect pre-trends.

**10.5 Random Effects / Hierarchical Models (If Sparse)**

Multilevel partial pooling across RUCC groups can stabilize noisy rural county estimates.

**10.6 Nonlinear & Threshold Models**

* Piecewise: remote effect kicks in only after RWct>0.25RW\_{ct} > 0.25.
* Quantile regression: remote work may affect distribution of LFPR differently across counties.

**11. Estimation Issues & Robustness**

**Measurement Error in Population Denominator:** Propagates into LFPR. Consider errors-in-variables sensitivity by perturbing POP16ctPOP16\_{ct} ± margin-of-error.  
**Small Sample Remote Estimates:** Use shrinkage to RUCC group means or Bayesian partial pooling.  
**Endogeneity:** Remote work adoption may respond to labor supply constraints (reverse causality). Instruments: broadband rollout, pre-pandemic industry telework capacity, state policy shocks.  
**Autocorrelation:** With 4 years, limited, but cluster at county.  
**Influential Observations:** Very small counties with volatile survey remote rates—winsorize or weight by population.

**12. Modeling Sequence Recommendation**

**Phase 1 – Descriptive:**

1. Plot LFPR by rural vs. urban over time.
2. Plot remote share by rural vs. urban.
3. Scatter: change in LFPR vs. change in remote share.

**Phase 2 – Baseline Regression:**  
4. FE regression: LFPR ~ Remote × Rural + Year FE + County FE.

**Phase 3 – Controls & Robustness:**  
5. Add broadband, industry mix, education.  
6. Re-estimate using alternative 16+ population methods (fixed ACS baseline).  
7. Use DiD high- vs. low-remote counties.

**Phase 4 – Interpretation:**  
8. Marginal effects of remote on LFPR in rural vs. urban (margins at means).  
9. Policy counterfactual: If rural remote rose to urban level, what LFPR change implied?

**13. R Implementation Skeleton**

Below is a streamlined pseudo-pipeline. Replace file paths with your own; API pulls optional.

# ---- Packages ----

library(tidyverse)

library(readr)

library(janitor)

library(lubridate)

library(fixest) # or plm / lfe

# ---- 1. Load LAUS: Annual county labor force ----

laus <- read\_csv("data/laus\_iowa\_annual\_2021\_2024.csv") %>%

clean\_names() %>%

select(county\_fips, year, labor\_force, employed, unemployed)

# ---- 2. Load ACS 2021 16+ baseline ----

acs16 <- read\_csv("data/acs2021\_pop16plus.csv") %>%

clean\_names() %>%

select(county\_fips, pop16plus\_2021 = pop16plus)

# ---- 3. Load PEP total population 2021-2024 ----

pep <- read\_csv("data/pep\_iowa\_county\_2021\_2024.csv") %>%

clean\_names() # cols: county\_fips, year, pop\_total

# ---- 4. Compute CAGR per county ----

pep\_wide <- pep %>%

pivot\_wider(names\_from = year, values\_from = pop\_total, names\_prefix = "pop\_")

pep\_growth <- pep\_wide %>%

mutate(

g = (pop\_2024 / pop\_2021)^(1/3) - 1

) %>%

select(county\_fips, g, pop\_2021, pop\_2024)

# ---- 5. Expand growth factors by year ----

years <- tibble(year = 2021:2024)

pop\_proj <- pep\_growth %>%

crossing(years) %>%

mutate(gf = (1 + g)^(year - 2021)) %>%

left\_join(acs16, by = "county\_fips") %>%

mutate(pop16plus = pop16plus\_2021 \* gf) %>%

select(county\_fips, year, pop16plus)

# ---- 6. Merge LF and pop16plus; compute LFPR ----

lfpr\_data <- laus %>%

left\_join(pop\_proj, by = c("county\_fips","year")) %>%

mutate(lfpr = 100 \* labor\_force / pop16plus)

# ---- 7. Remote work (survey aggregated to county-year) ----

remote <- read\_csv("data/remote\_share\_county\_year.csv") %>%

clean\_names() %>%

select(county\_fips, year, remote\_share)

# ---- 8. Rural classification (RUCC) ----

rucc <- read\_csv("data/rucc\_codes\_iowa.csv") %>%

clean\_names() %>%

mutate(rural = if\_else(rucc\_code >= 4, 1, 0)) %>%

select(county\_fips, rural, rucc\_code)

# ---- 9. Build analysis panel ----

panel <- lfpr\_data %>%

left\_join(remote, by = c("county\_fips","year")) %>%

left\_join(rucc, by = "county\_fips")

# ---- 10. Add optional controls (broadband, industry, education) ----

# controls <- read\_csv("data/controls\_county\_year.csv")

# panel <- left\_join(panel, controls, by = c("county\_fips","year"))

# ---- 11. Descriptive grouped means ----

panel %>%

group\_by(rural, year) %>%

summarise(mean\_lfpr = mean(lfpr, na.rm = TRUE),

mean\_remote = mean(remote\_share, na.rm = TRUE),

n = n())

# ---- 12. Fixed effects regression ----

# Model: LFPR ~ Remote + Remote\*Rural + Year FE + County FE

model\_fe <- feols(lfpr ~ remote\_share \* rural | county\_fips + year, data = panel, cluster = ~ county\_fips)

summary(model\_fe)

# ---- 13. Marginal effects ----

etable(model\_fe)

# Predict effect of +10pp remote in rural vs. urban

# ---- 14. Save panel ----

write\_csv(panel, "output/iowa\_county\_lfpr\_panel\_2021\_2024.csv")

**14. Data Quality Diagnostics**

Perform these checks before final analysis:

1. **Coverage:** Count counties per year with non-missing LF, pop16plus, remote.
2. **Extreme LFPR:** Flag >85% or <35%. Investigate denominators.
3. **Remote share sampling error:** Compute standard errors from survey; mark low-sample cells.
4. **Year-over-year jumps in population:** Compare PEP growth vs. ACS; cap extreme factors.
5. **Correlation checks:** LFPR vs. remote share—plot scatter, LOESS fit.

**15. Reporting & Visualization Ideas**

* **Choropleth maps:** LFPR by county each year; remote share overlay.
* **Slopegraph:** Rural vs. urban mean LFPR from 2021→2024.
* **Scatter w/ regression line:** Change in LFPR vs. change in remote share by county.
* **Coefficient plots:** Display interaction term estimates across model specs.

**16. Extensions & Sensitivity Analyses**

**Population denominator sensitivity:** Recompute LFPR using fixed 2021 ACS baseline only; compare coefficients.  
**Alternative rural coding:** RUCC binary vs. RUCC 3-level vs. Metro/Nonmetro.  
**Lag structure:** Use lagged remote share (prior year) to reduce simultaneity.  
**Instrumental variable:** Broadband expansion or pre-pandemic telework capacity index as IV for remote adoption.  
**Weighting:** Population-weight regressions vs. unweighted county counts.

**17. Reproducibility & Documentation Checklist**

* Document original sources (download date, version).
* Store raw data in /data/raw/; processed in /data/clean/.
* Maintain data dictionary with variable definitions & units.
* Track ACS margins of error; record treatment of them.
* Scripted pipeline (R project / targets / renv).
* Version control (Git) with commit logs for data updates.
* Output tables reproducible from code.

**Final Notes**

This document captures the working blueprint for constructing an Iowa county-year panel to study the remote work–LFPR relationship in rural vs. urban settings for 2021–2024. As data are assembled, we can iteratively refine: (1) the population scaling method; (2) remote definitions; and (3) model specifications.

**Let me know when you’re ready to move into coding, diagnostic plots, or model estimation.**