LEAP Series, SERD ADB
May 18, 2021

# From Battlefield to Marketplace:

## Connectivity, Industrialization, and Spatial Convergence in the Greater Mekong Sub-Region

Manabu Nose (IMF), Yasuyuki Sawada (ADB),
and Tung Nguyen (Hitotsubashi U)

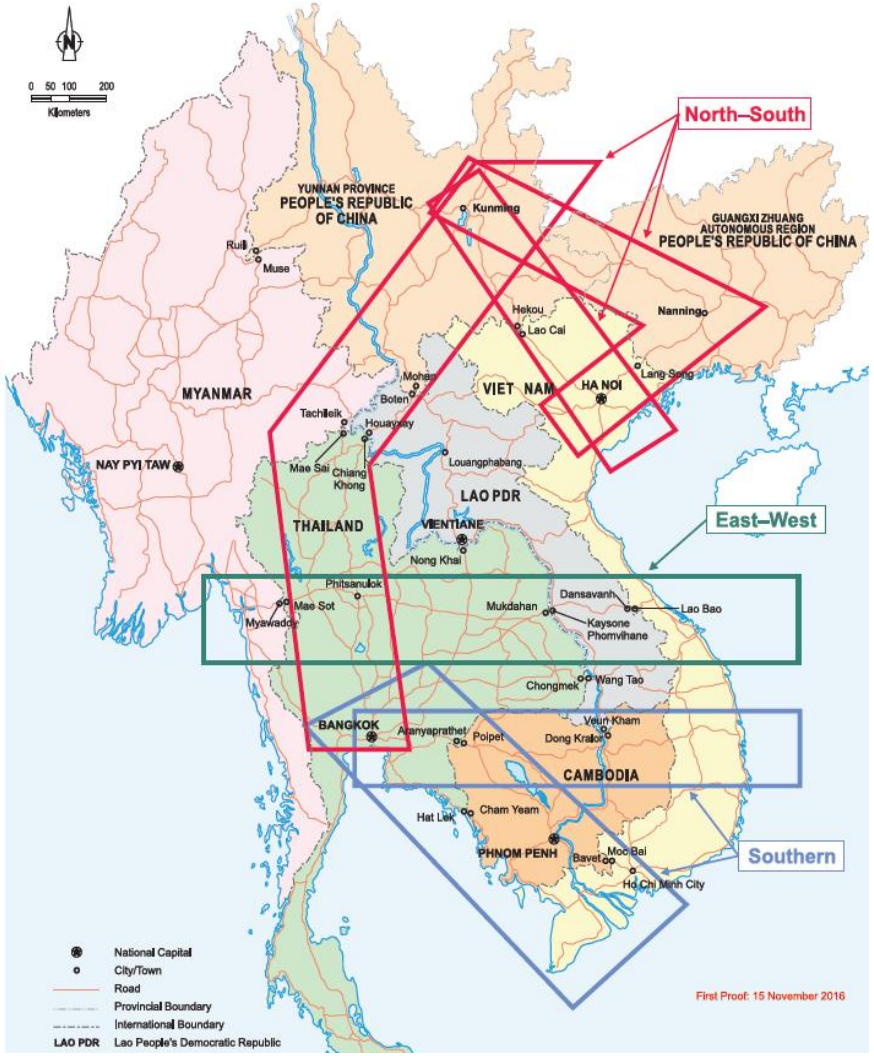
# Questions

- Would it be possible to conduct rigorous IE of ADB's flagship GMS Program?
  - IE of a large scale infra project
- Overall gains in job creation, industrialization, and fiscal multipliers?
- Reduces or exacerbates (spatial) inequality?

# What we do?

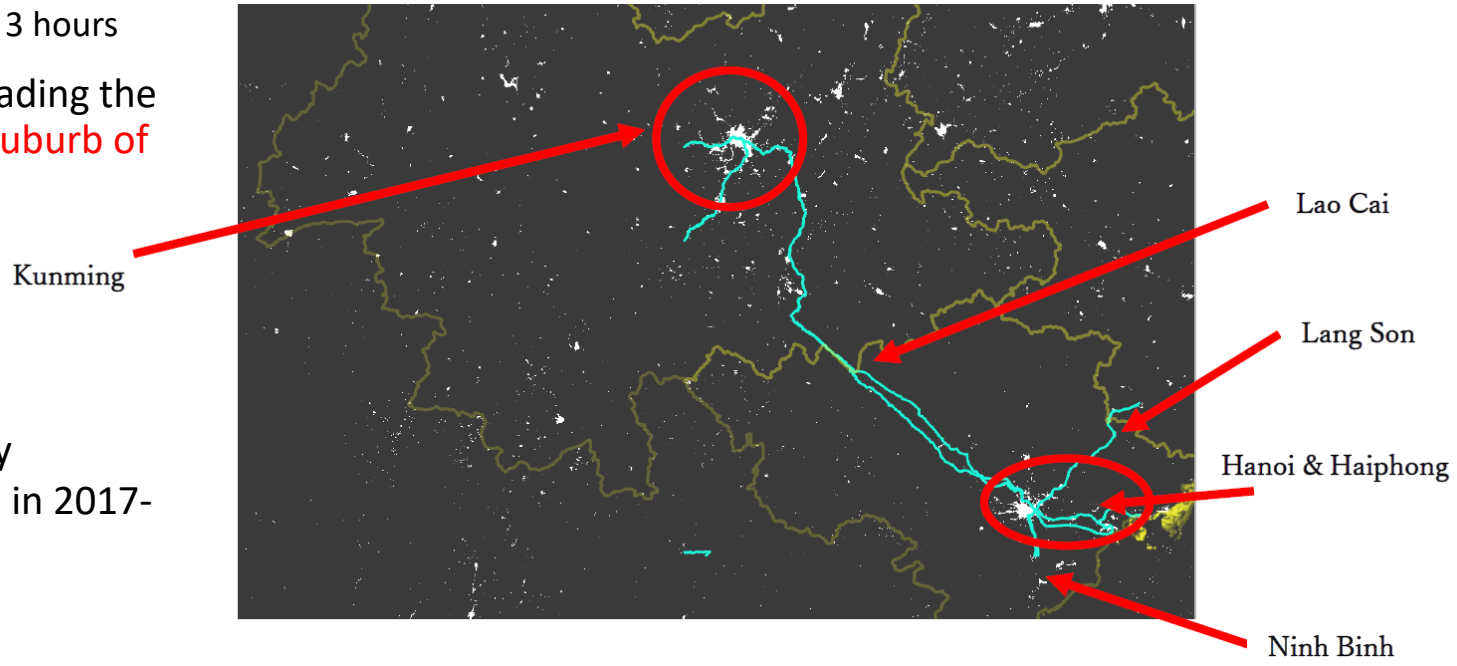
- IE of transport projects in the northern part of Vietnam: GMS Kunming–Hanoi, Haiphong, and Hanoi–Lang Son corridors
- Employ satellite imageries and ground-level data to grasp overall trends
- With firm and HH data, estimate the nexus among "interregional highways – market potential – industrial development"
  - Production, entry, and employment

# GMS North-South Corridor

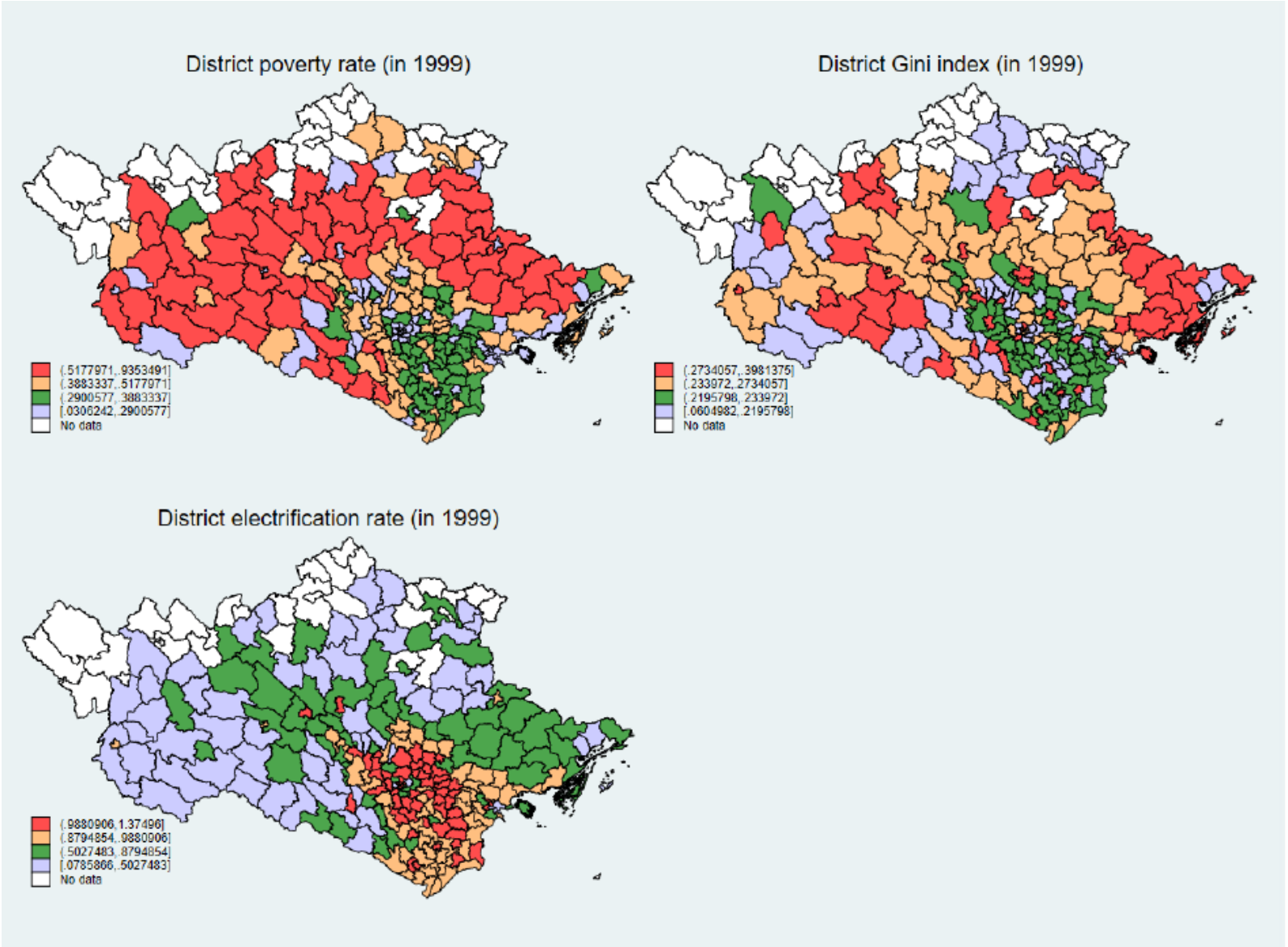



# GMS highways in Northern Vietnam

- The Hanoi – Lao Cai (HNLC) expressway: (financed by ADB)
  - 1.4 bil. USD invested on 244km expressway, constructed from 2009-2014.
  - Reduction in travel time from at least 10 hours to about 3 hours
- In parallel, 160 million USD was invested on upgrading the 285 km railway from Yen Vien station (northern suburb of Hanoi) to Lao Cai.
- Hanoi – Haiphong (HNHP) expressway:
  - 2.4 bil. USD invested through BOT scheme
  - Complete the Kunming-Haiphong corridor
- Hanoi – Ninh Binh expressway: upgraded & newly constructed, completed in 2014 (further upgrade in 2017-2018)
- Hanoi – Lang Son (HNLS) expressway: 3 sections
  - Hanoi-Bac Giang (50.9km) completed in 2016
  - Bac Giang-Chi Lang (63.1km, Lang Son) completed in 2019
  - Chi lang – Huu Nghi (43.3km, Lang Son): initially planned to be financed by ADB's loan, but stalled



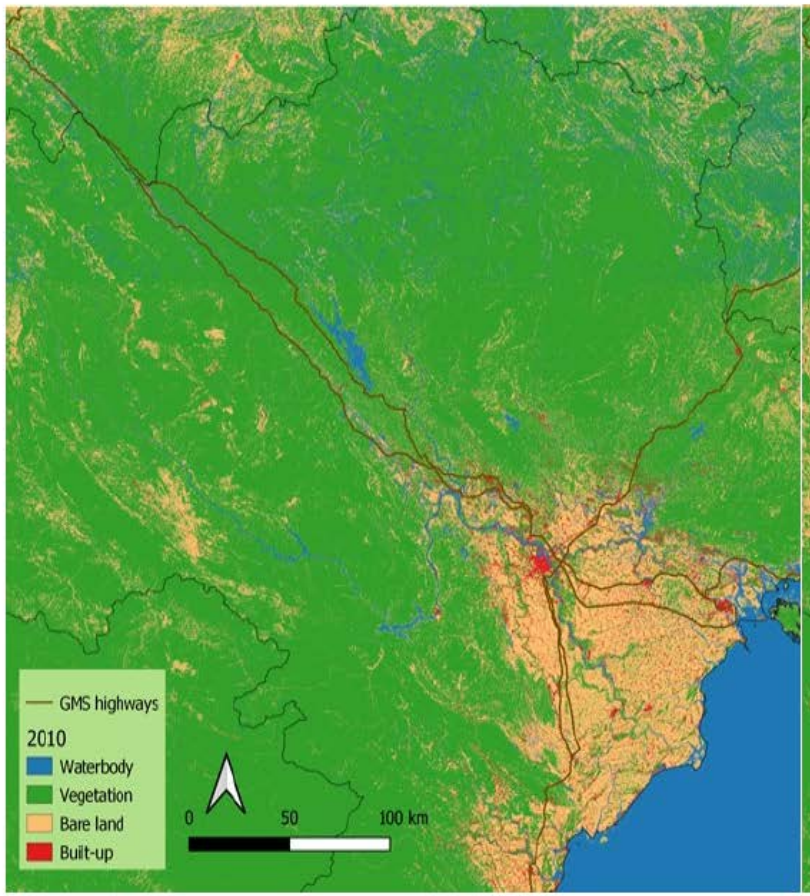
# Baseline conditions: “Core & Periphery” pattern



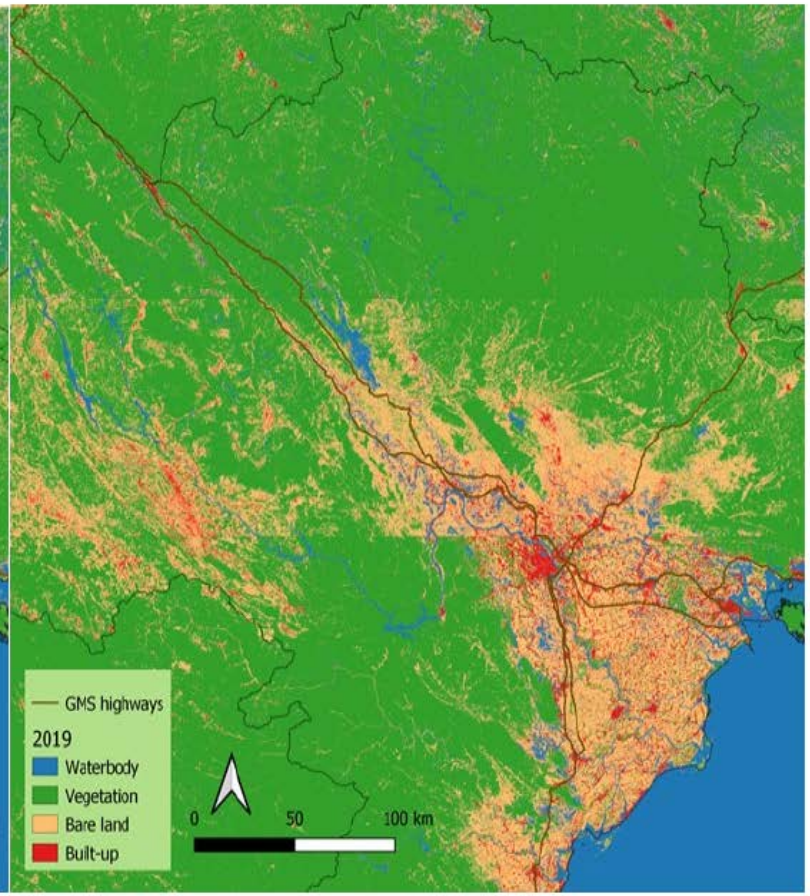
# Baseline conditions and development

## Landcover Mapping: Landsat Images

2010



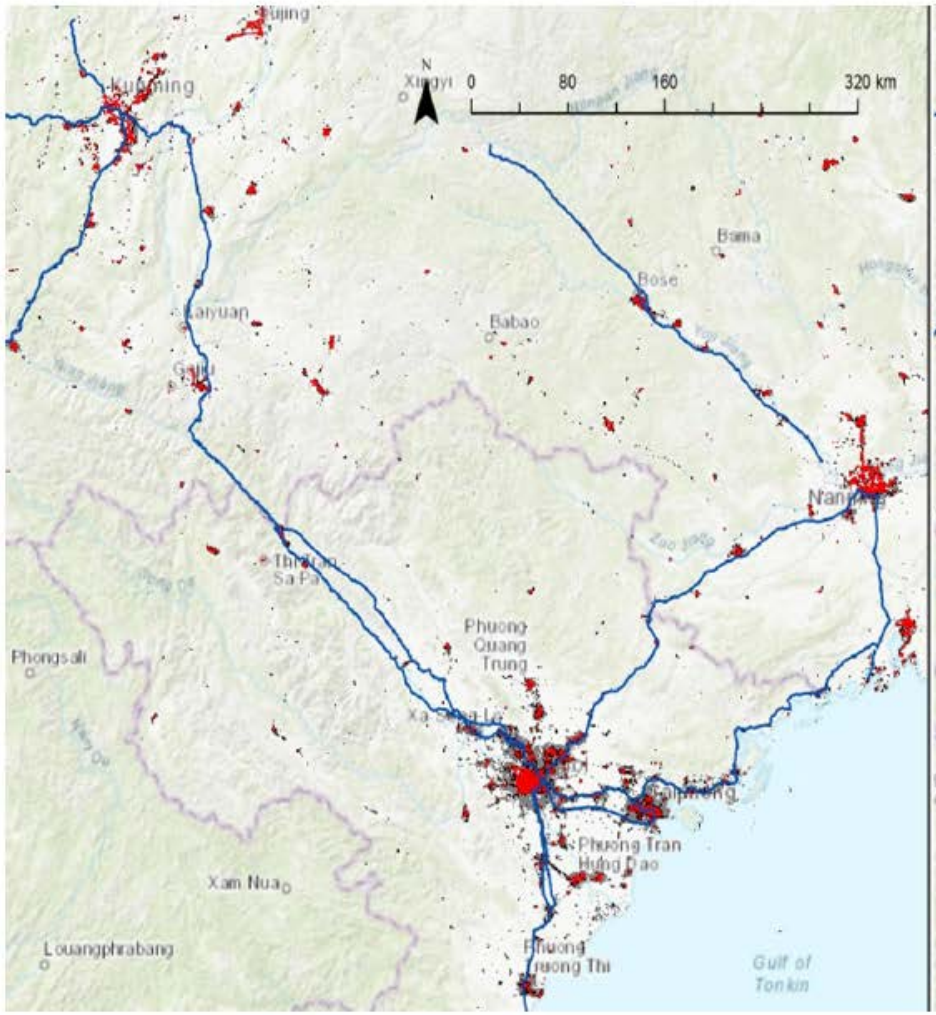

2019



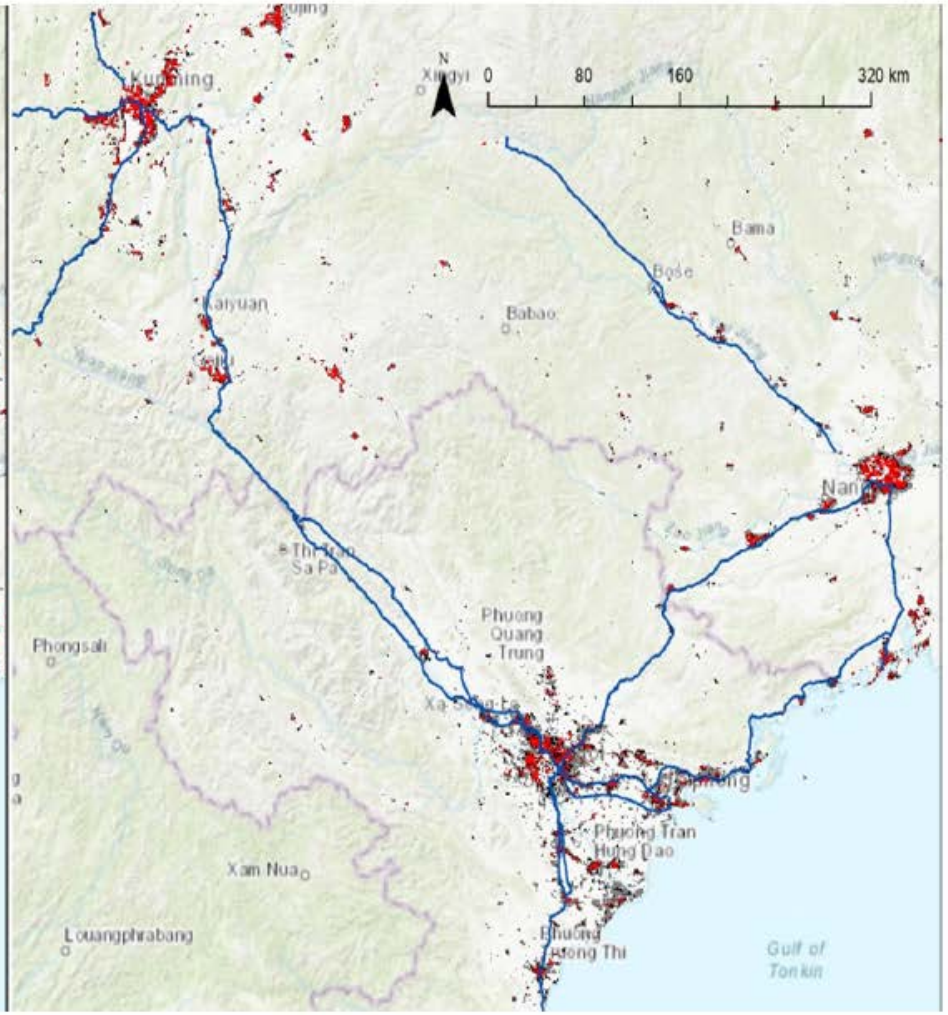

Note: supervised classification using neural network algorithm is applied to training data from Landsat 5 and 8 images.

# Baseline conditions and development

NTL difference, from 2012-2016




NTL difference, from 2016-2019



# Research design

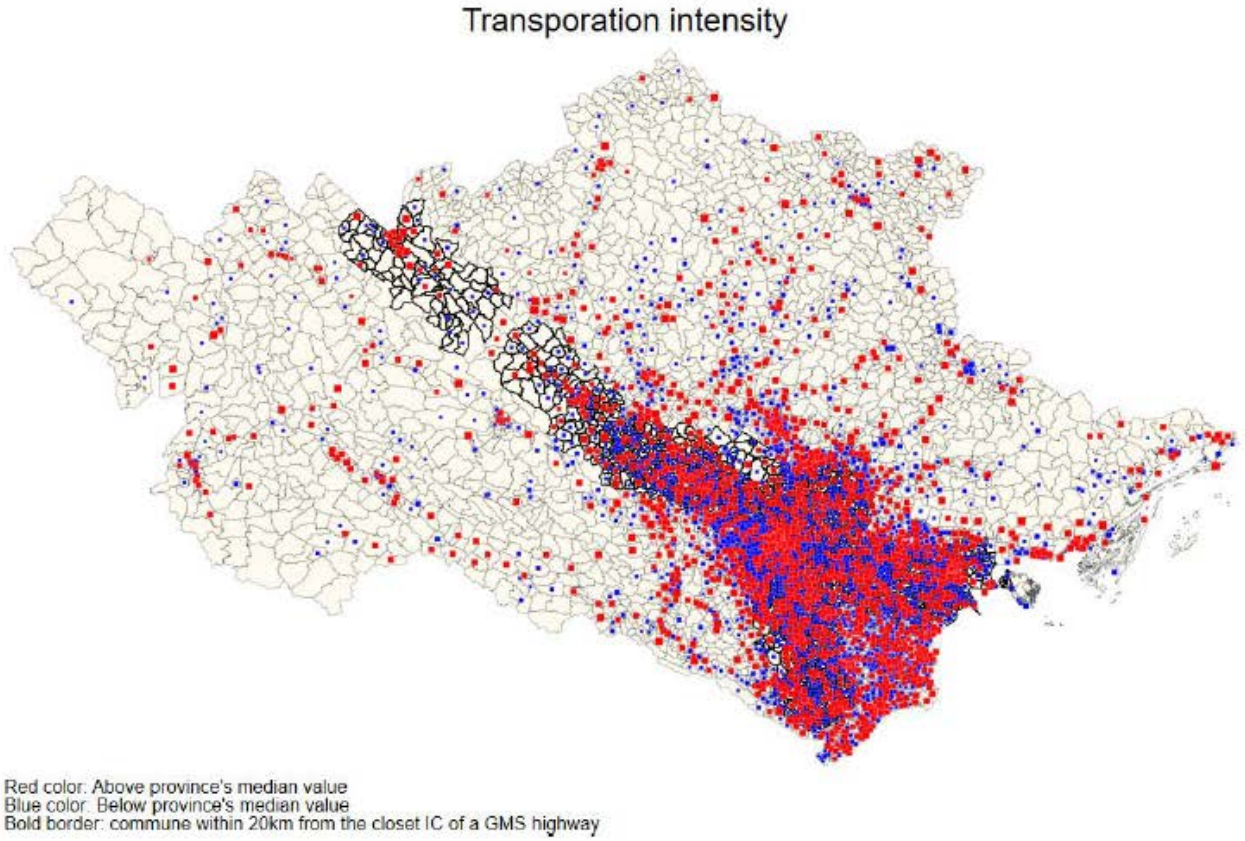
- How can we attribute the observed development patterns from NTL and build ups to GMS?

- GMS highway construction as exogenous treatments on structural change and employment generation

- Methods:
  - DID (with spillover effects)
  - IV regression

# Impacts of GMS

- The transportation reliance based input-output (IO) table
- Using the detailed GSO's IO table covering 138 sectors, we compute the average transport intensity of each commune using the initial share of each industry's output as the weight.
- We classify each commune into treatment group if the commune's transport intensity measure is greater than the threshold to the province's median or 75th percentile values.

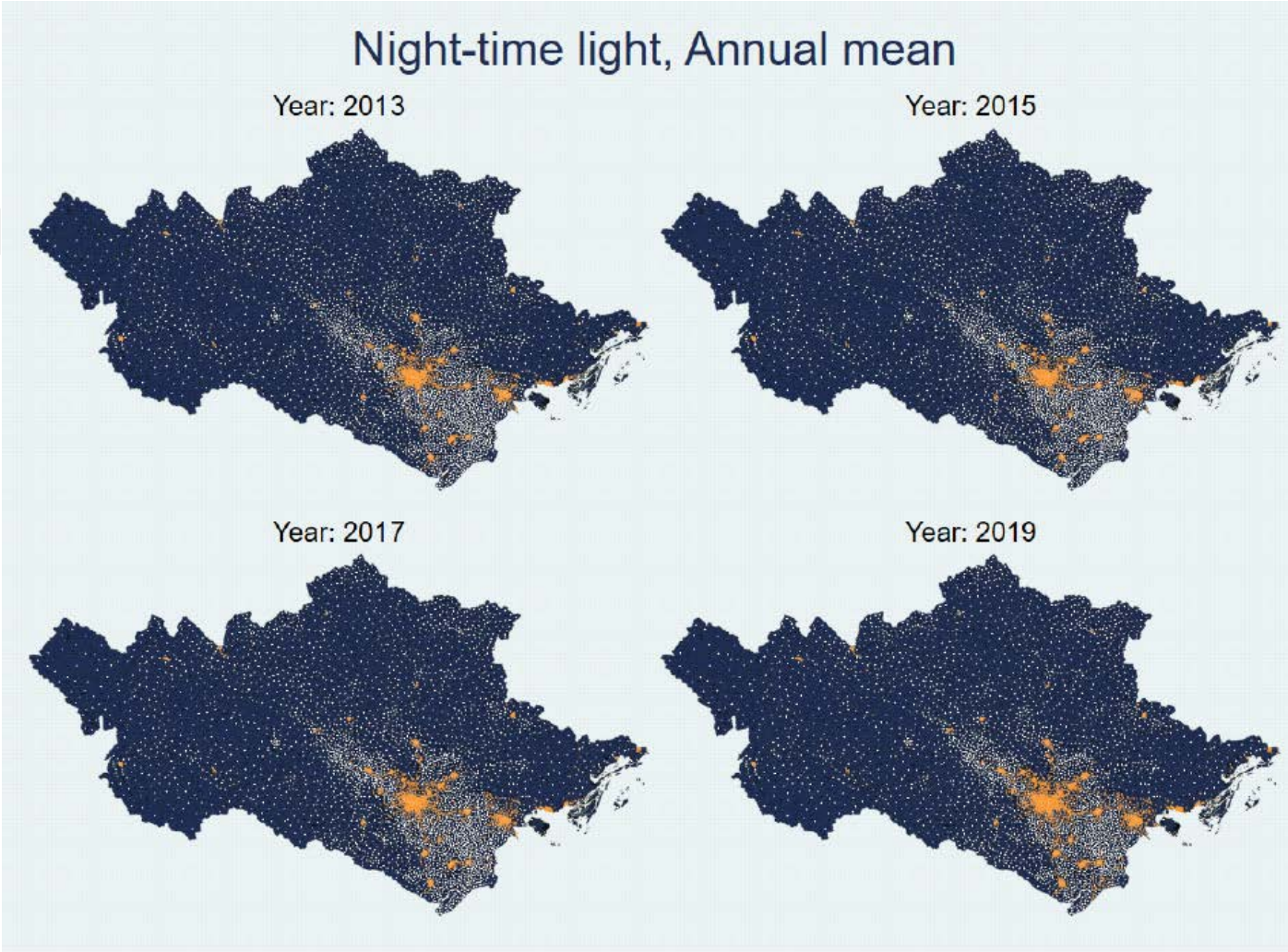



Source: National input-output table in 2011, VEC

Figure 13: Spatial Variation in Transportation Intensity

# Impacts of GMS on NTL (VIIRS)

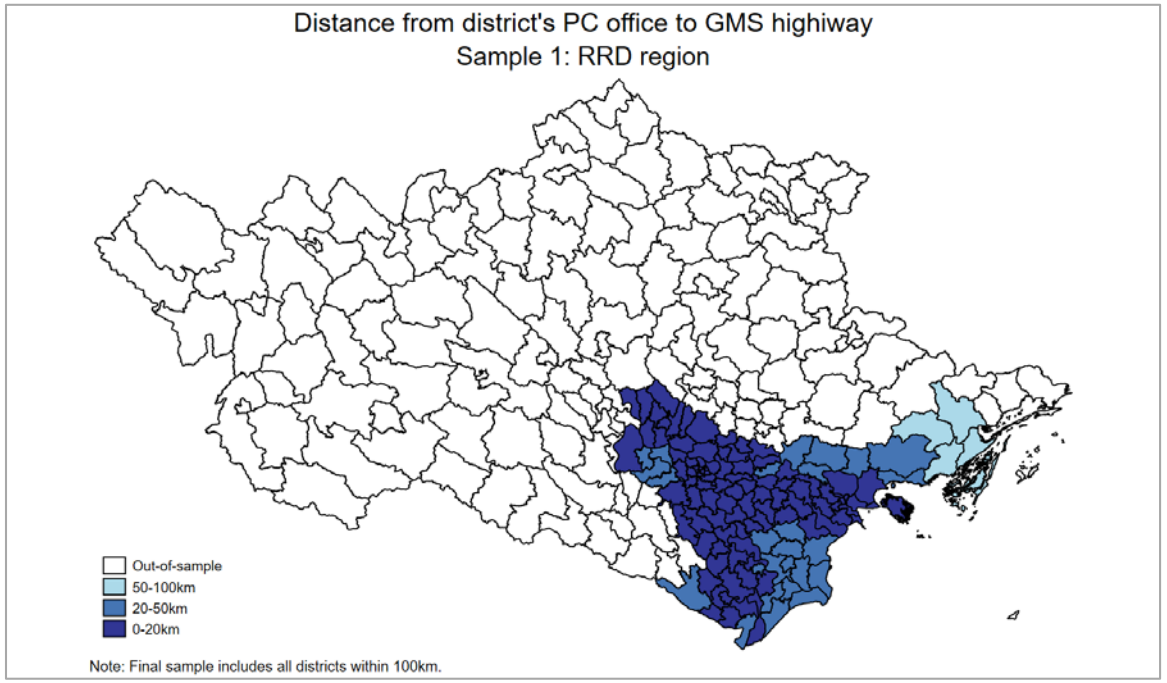
- DID
  - *d*=1[w/i 20km from GMS highways & transportation intensity > p50]
- 2012-2016
  - RRD (+66.3 points)
  - NE/NW (neutral)
- 2012-2019
  - RRD (+97.2 points)
  - NE/NW (+71.6 points)

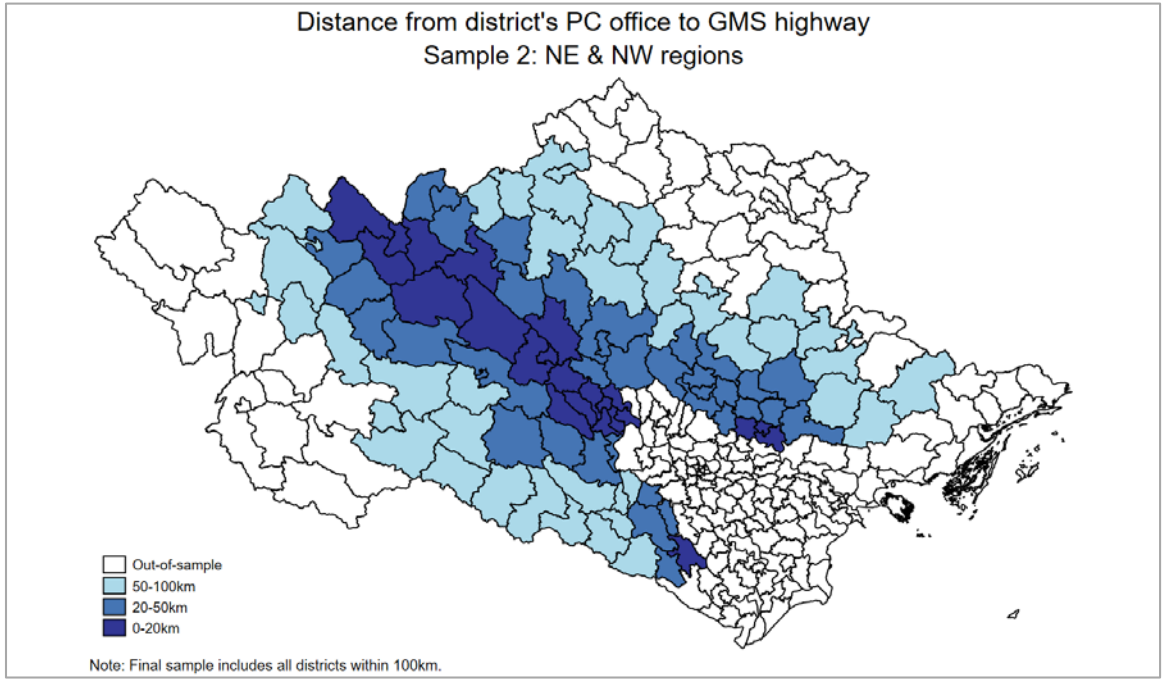


# Ground-level data

- District or commune panel data from GSO
  - Firm Census (2011 & 2016), Annual Surveys (VES: 2006-2016)
    - Wage, employment, and firm financial data
  - Vietnam Household Living Standard Survey (VHLSS)
    - Population and housing price
- Restricted to districts within 100km from the GMS highway

| | Total | 0-20km | 20-50km | 50-100km |
|---|---|---|---|---|
| | Number of districts | | | |
| RRD region | 124 | 94 | 26 | 4 |
| NE & NW regions | 101 | 23 | 36 | 42 |
| | Number of communes | | | |
| RRD region | 2,329 | 1,563 | 714 | 52 |
| NE & NW regions | 1,225 | 383 | 502 | 340 |



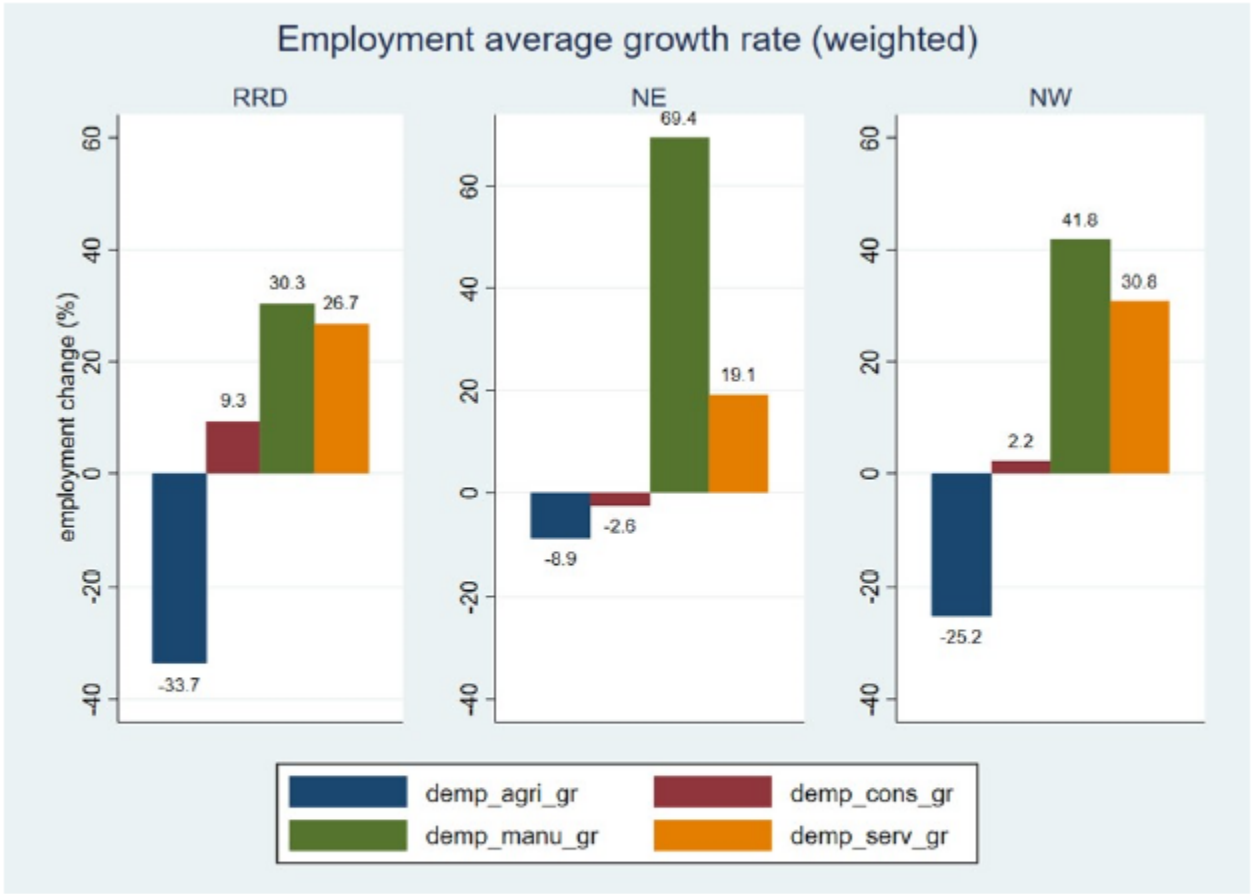
# Reconfirm overall development

| | Population | | |
|---|---|---|---|
| Regions | Population in 2009 | Pop. GR 1999-2009 | Pop. GR 2009-18 |
| Red River Delta | 152,956 | 10% | 22% |
| North East | 86,557 | 5% | 35% |
| North West | 69,193 | 19% | 48% |

| | Enterprises | | | |
|---|---|---|---|---|
| Regions | # of Ent. in 2011 | # of Ent. GR 2011-16 | o/w Manufacturing | o/w Service |
| Red River Delta | 104,022 | 51% | 46% | 52% |
| North East | 9,011 | 38% | 41% | 48% |
| North West | 5,013 | 43% | 26% | 55% |

Source: Population Census 1999 & 2009; VHLSS 2018; VES 2011 & 2016

# Reconfirm overall development and st change




Source: district-level VES data. Note: employment growth rate is calculated using the 2011 employment level as the weight.

# Impacts of GMS on firms and employment

- DID for 5-yr impacts (2011-2016) of *d*=1[within 20km from GMS & transportation intensity > p75]

- RRD
  - # of firms: Manufacturing (+7.5%) and service (-10.9% to -4.8%)
  - Employment: Manufacturing (+28.1%) and service (-23.3% to -13.8%)
  - Employment share: Manufacturing (+8.0%) and service (-10.1% to -4.1%)

- NE/NW
  - # of firms: Manufacturing (+12.8%) and service (-18.4% to neutral)
  - Employment: Manufacturing (+83.8%) and service (neutral)
  - Employment share: Manufacturing (17.8%) and service (-10.1% to neutral)

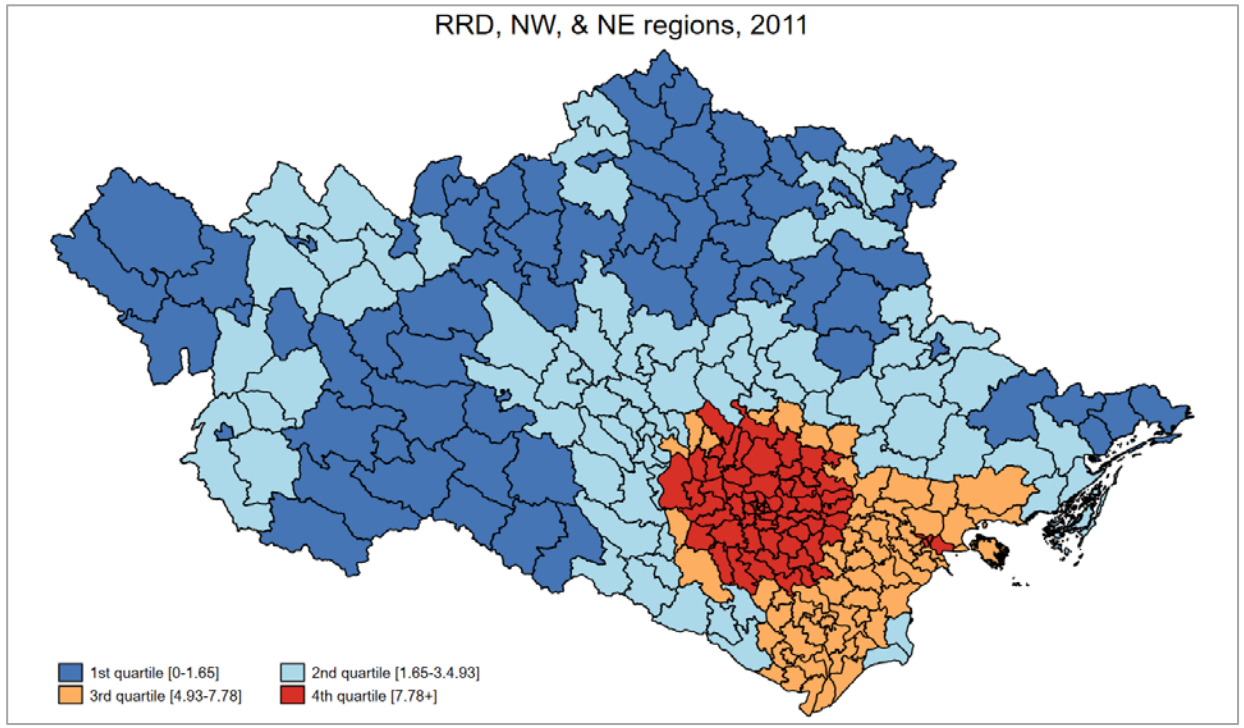
# Mechanisms?

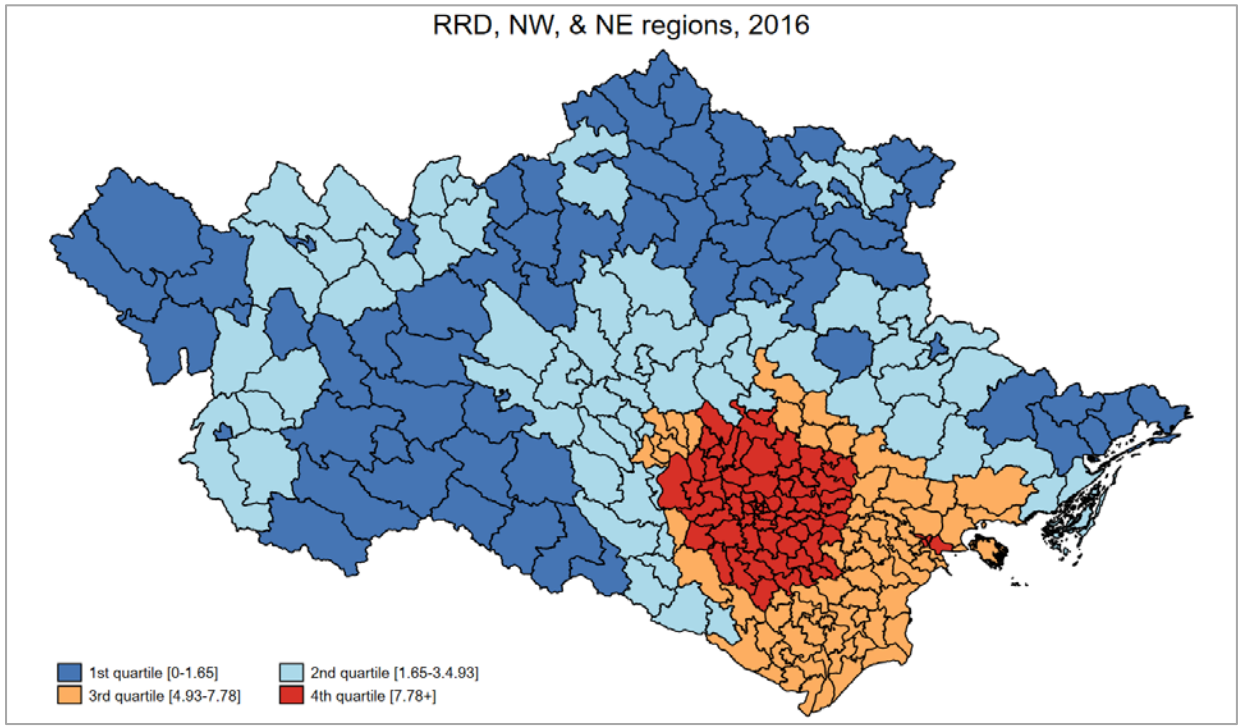
- ❑Availability of labor
- ❑Existence of consumer
- ❑Market size

- ❑The canonical Krugman or Fujita-Krugman-Venables model:
  - ❑Market size ($\pi$)
  - ❑Increasing returns scale ($\sigma^{-1}$)
  - ❑Transportation cost ($\tau$)

# Mechanisms? Harris MP to capture agglomeration

- $MP_{ct} = \sum_{l \neq c} \frac{emp.density_{lt}}{d_{cl}}$

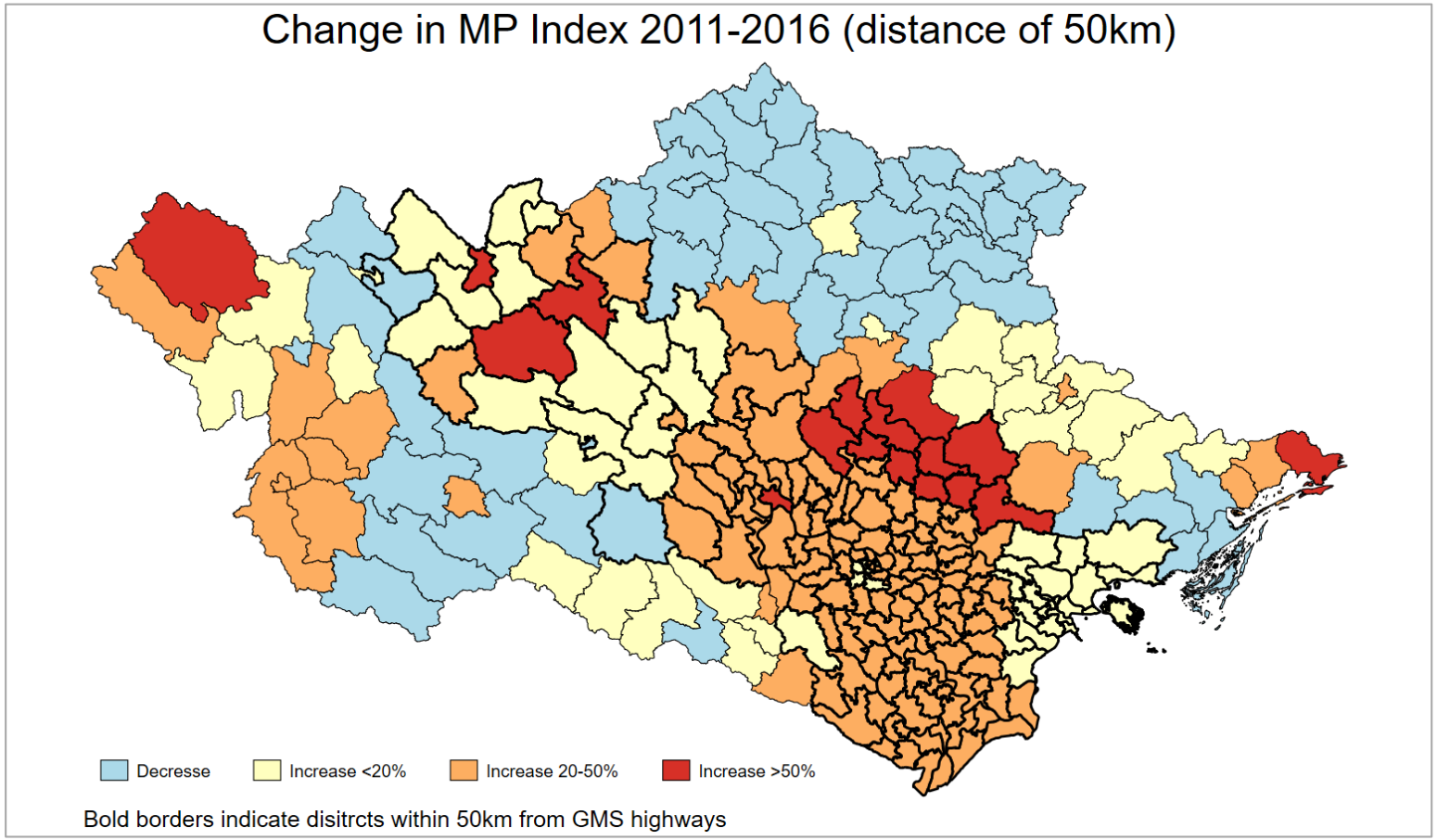



Note: MP index is estimated using total district employment from the VES (registered firms) and the distance of 50km between two district People Committee's Offices. Category based on quartiles of the rounds 2011 and 2016 combine.

# Mechanisms? Harris MP to capture agglomeration

- The MP index concentrated around RDD and has expanded over time
- Faster increases in MP index in districts near the GMS highways
- Districts in NE region seems to be benefited more from GMS highways due to location advantage

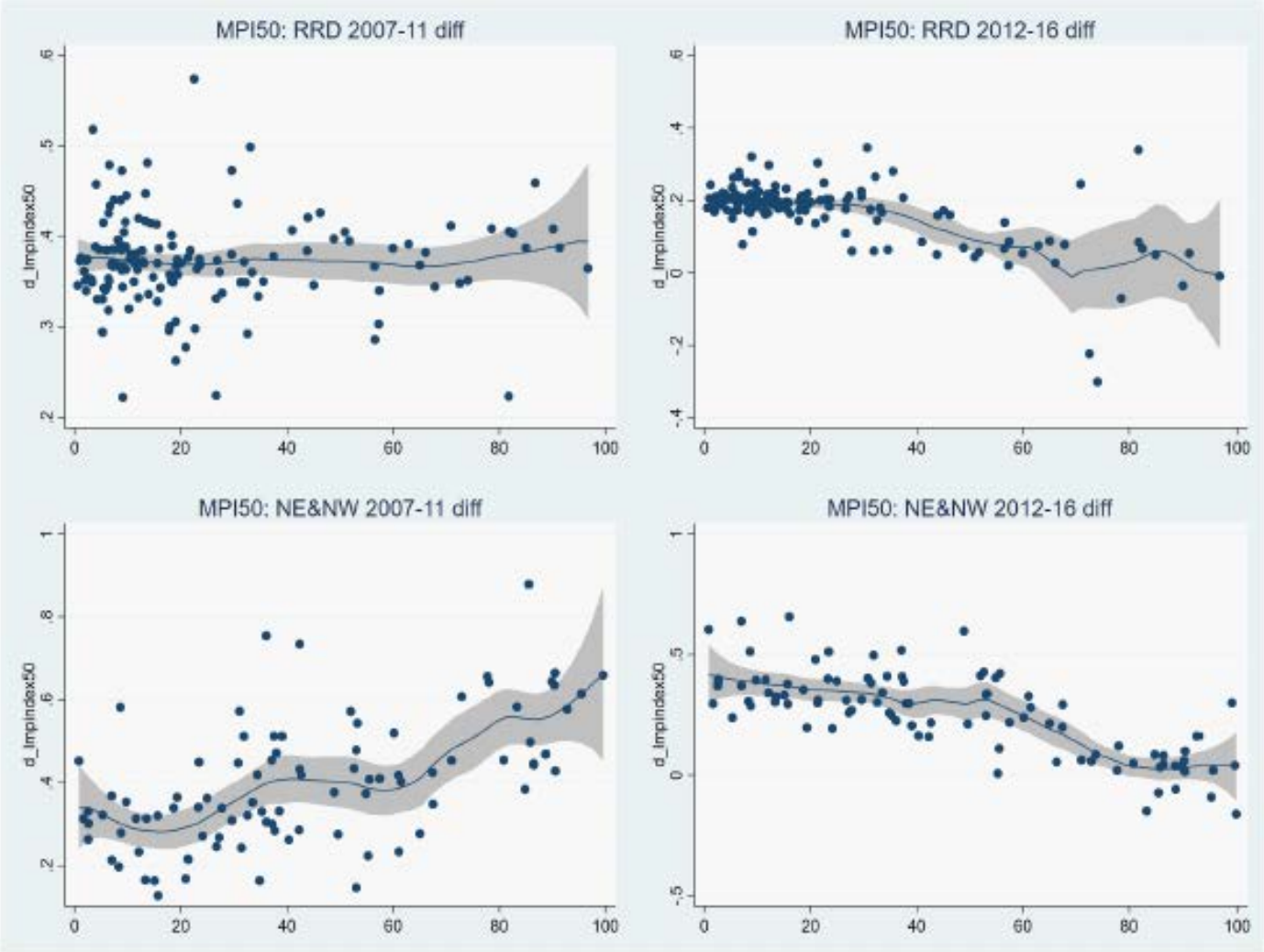



Note: MP index is estimated using total district employment from the VES (registered firms) and the distance of 50km between two district People Committee's Offices..

# GMS impact on market potential (MP vs distance)

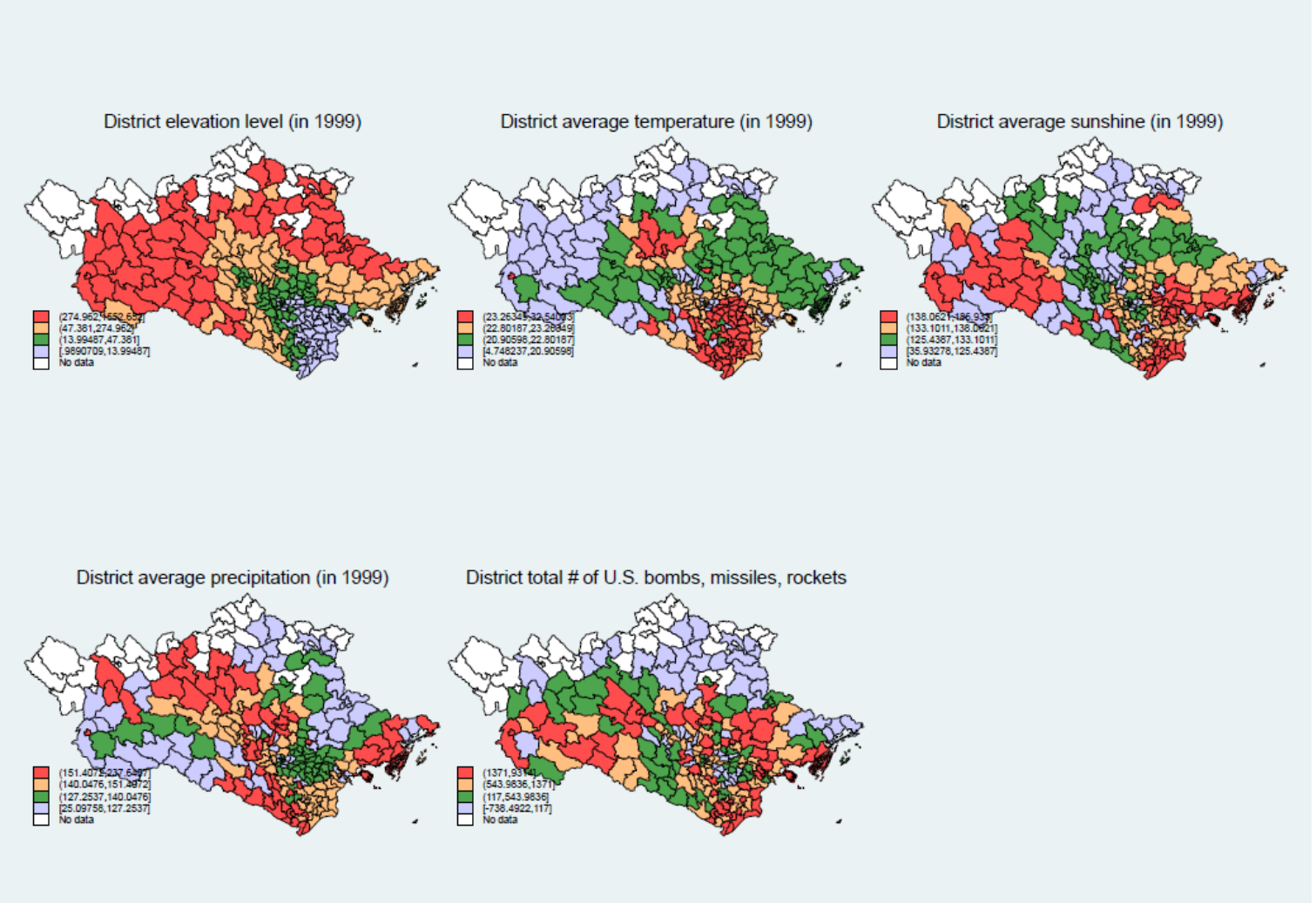
**Before** GMS

**After** GMS



# Determinants of MP, given distance to GMS?

- RRD
  - Temperature (+)
  - Sunshine (-)
  - Precipitation (-)
- NE/NW
  - Elevation (-)
  - US bomb (+)



# Impacts of MP on firm agglomeration

- 1% increase in MP results in:

- RRD
  - # of firms in manufacturing (+0.81%) and service (+0.74%)
  - (55%, 60%, and 49% increase in # of small, medium, and large-scale enterprises)
  - Employment share in manufacturing (neutral) and service (neutral)

- NE/NW
  - # of firms in manufacturing (+0.83%) and service (neutral)
  - (66% increase in # of small-scale enterprises)
  - Employment share in manufacturing (+0.35%) and service (neutral)

# Remarks

- Initial C-P pattern has been transformed by GMS
  - Interregional highways facilitated the economy-wide, large-scale industrialization and significant income convergence across space.
  - 1% increase in market potential results in about 0.8% increase in number of manufacturing firms in both core and peripheral cities
  - Impacts are more salient in the peripheral cities.

- Heterogenous impacts between GMS “treatment” areas and “control” areas
  - Agglomeration economies and the shifts to manufacturing industries in the area near the highways.
  - Also, triggering labor reallocation from agriculture to manufacturing in rural peripheries.
  - Local MP significantly expanded within 20km from the highway

# Remarks (continued)

- Use of big data, national statistics, and administrative data for IE
  - High-resolution light-at-night data and Landsat aerial images
  - Firm/HH microdata
  - Higher frequency mobility data (mobile phones and satellites)

| Theme | Source | Observation frequency | Available frequency for cloud-free composite | Latency | Observation period |
|---|---|---|---|---|---|
| **Land cover & Land use** | Landsat 1-3 MSS | 16 days, but differ by region. | Yaerly but not available for some years. | NA | 1972 - 1983 |
| | Landsat 5 TM | 16 days, but differ by region. | Yearly but not available for some years. | NA | 1984 -2013 |
| | Landsat 8 OLI | 16 days, quite few missing. | Yearly | Near real-time (1 day) | 2013 - present |
| **Night-time light** | VIIRS | Everyday | Monthly | Near real-time (1 day) | 2012 - |
| | DMSP-OLS | Everyday | Yearly | NA | 1992 - 2013 |

# Future Plans

- JFPR “Quality of Infrastructure” TA
  - KSTA 54114-001: Using Frontier Technology and Big Data Analytics for Smart Infrastructure Facility Planning and Monitoring (led by Dr. Takashi Yamano, EREA)
- Planned GMS studies under this TA
  - IE of Vietnam and China trade corridors
  - Large scale IE of GMS using structural trade models and long-term satellite imageries
  - IE of GMS using recent mobility data

## Main Satellite Data Sources Used in Economics

| *Source* | *Economics applications* | *Highest resolution* | *Pricing* | *Availability by year* | *Examples* | *For more information* |
|---|---|---|---|---|---|---|
| Landsat | Urban land cover, beaches, forest cover, mineral deposits | 30 m | Free | 1972– (8 satellites) | Foster and Rosenzweig (2003); Faber and Gaubert (2015) | http://landsat.usgs.gov/ |
| MODIS | Airborne pollution, fish abundance | 250 m | Free | 1999– (Terra); 2002– (Aqua) | Foster, Gutierrez, and Kumar (2009); Burgess et al. (2012) | http://modis.gsfc.nasa.gov/ |
| Night lights (DMSP-OLS, VIIRS) | Income, electricity use | ~1 km | Free digital annual (DMSP- OLS) and monthly (VIIRS) composites | Digital archive 1992–2013+ (VIIRS 2012–; film archive 1972–1991) | Henderson et al. (2012); Chen and Nordhaus (2011) | http://ngdc.noaa.gov/eog/ |
| SRTM | Elevation, terrain roughness | 30 m | Free | 2000 (static) | Costinot, Donaldson, and Smith (2016) via Global Agro-Ecological Zones (GAEZ) data | https://lta.cr.usgs.gov/SRTM |
| DigitalGlobe (including Quickbird, Ikonos) | Urban land cover, forests | < 1 m | Not free | 1999– (6 satellites) | Marx, Stoker, and Suri (2015); Jayachandran, de Laat, Lambin, and Stanton (2016) | https://www.digitalglobe.com/resources/satellite-information |

Source) Table 1
Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives*, 30 (4): 171-98.

- Appendix

# Background

- Firms near GMS highways have significantly increased their passenger volume and freight volume compared to other firms.
- Freight transportation increase faster in area closer to highways (0-20km) while passenger transportation increase in wider area (0-50km).
- Matched firms close to the highway experience faster increase than newly established firms, but with lower growth in real revenue.
- Passenger transportation increase faster in RRD region while freight transportation increase faster in NE & NW regions.

Table 1. Passenger and Freight Transportation Growth 2011-2016

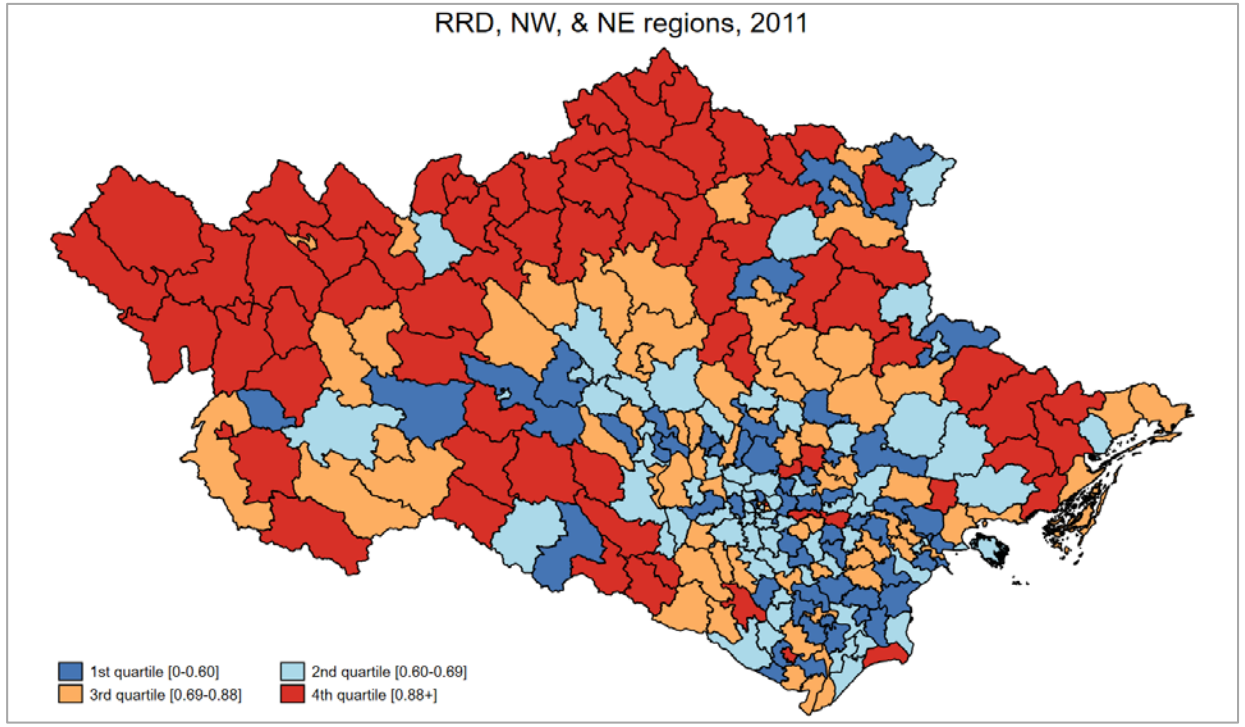
| | Distance to the neartest highway | Passenger transportation | | | Freight transportation | | |
|---|---|---|---|---|---|---|---|
| | | Total revenue (mil. VND) | Volume (people) | #of firms (2011/2016) | Total revenue (mil. VND) | Volume (tons) | #of firms (2011/2016) |
| | | ***Panel A: All recorded firms*** | | | | | |
| Red River Delta | 0-20km | 36.3% | 969.9% | 1066/1468 | 41.9% | 554.2% | 2541/4258 |
| | 20-50km | 30.9% | 931.3% | 180/206 | 23.7% | -62.5% | 451/518 |
| | 50-100km | 46.7% | -3.0% | 24/31 | -25.9% | 136.5% | 74/100 |
| North East & North West | 0-20km | 41.3% | 55.7% | 120/140 | 149.7% | 294.3% | 317/546 |
| | 20-50km | 46.9% | 124.9% | 96/111 | 47.1% | -94.5% | 376/493 |
| | 50-100km | 79.0% | -93.2% | 73/93 | 73.8% | -91.2% | 206/256 |
| | | ***Panel B: Matched firms only*** | | | | | |
| Red River Delta | 0-20km | 3.6% | 1308.4% | 682/682 | 18.3% | 782.2% | 1440/1440 |
| | 20-50km | 4.5% | 1227.5% | 100/100 | -27.2% | -83.8% | 141/141 |
| | 50-100km | 39.5% | 15.7% | 16/16 | -15.5% | 155.7% | 31/31 |
| North East & North West | 0-20km | 21.4% | 56.0% | 73/73 | 17.9% | 841.2% | 121/121 |
| | 20-50km | 27.8% | 62.3% | 57/57 | 13.5% | -96.3% | 171/171 |
| | 50-100km | 23.1% | -93.8% | 59/59 | 29.3% | -79.5% | 86/86 |

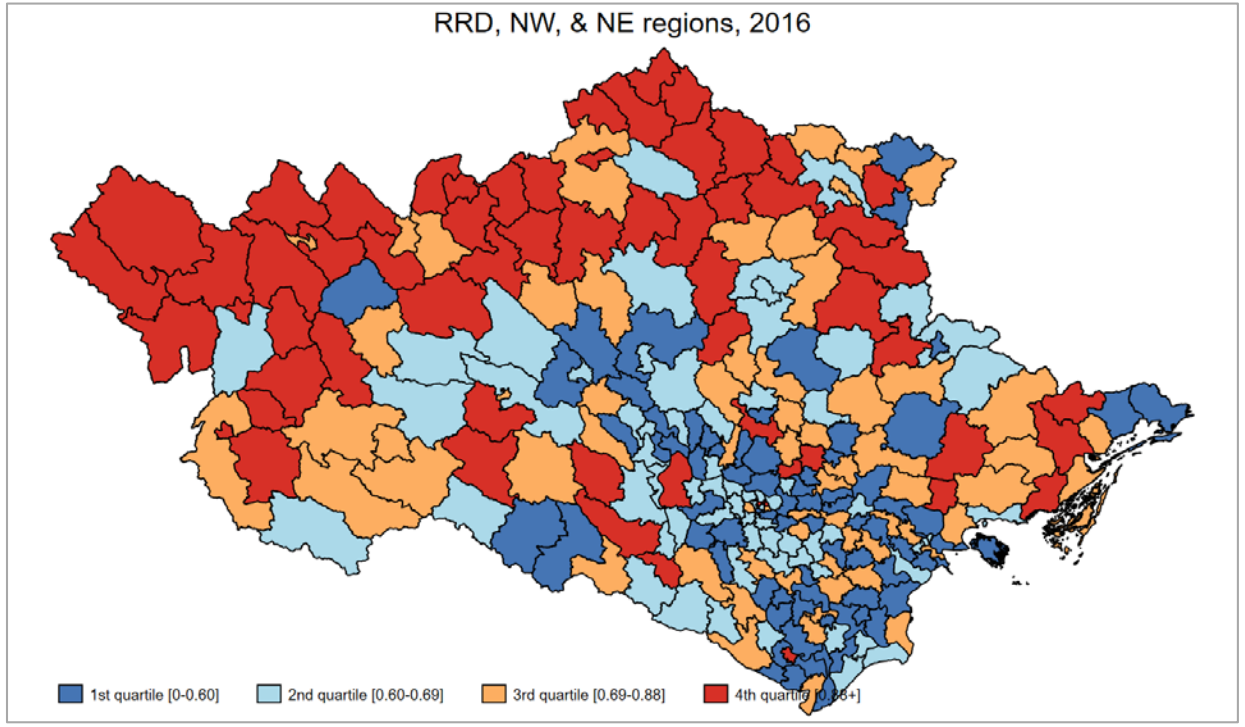
Note: Data is aggregated at the region-distance level, using firm-level enterprise censuses 2011 and 2016. Only registered firms are accounted for. Unit for each original variable are in parentheses. About 25% of firms exist in both censuses. Distance to the nearest highway based on firm location at the commune level.

# Agglomeration pattern: Krugman Index

$$K - index_{ct} = \sum_{s} \left| \frac{emp_{cst}}{emp_{ct}} - \frac{emp_{st}}{emp_{t}} \right|,$$

where $emp_{cst}$ is the employment share of industry $s$ (agriculture, manufacturing, construction, service) in district $c$ at year $t$. Market in district $c$ is more specialized as *K-index* increases.

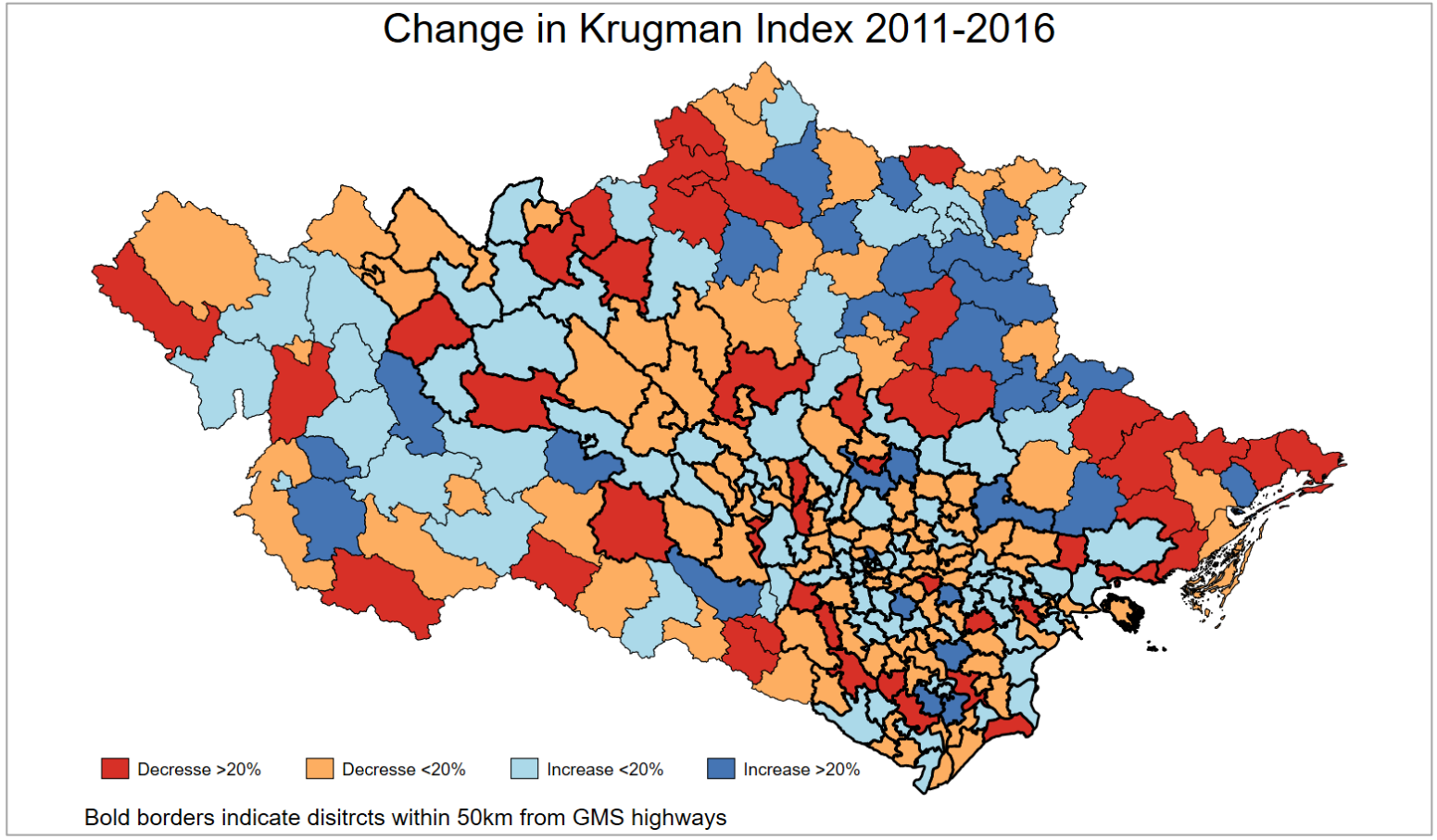



Note: Krugman index is estimated using total district employment from the VES (registered firms). Category based on quartiles of the rounds 2011 and 2016 combine.

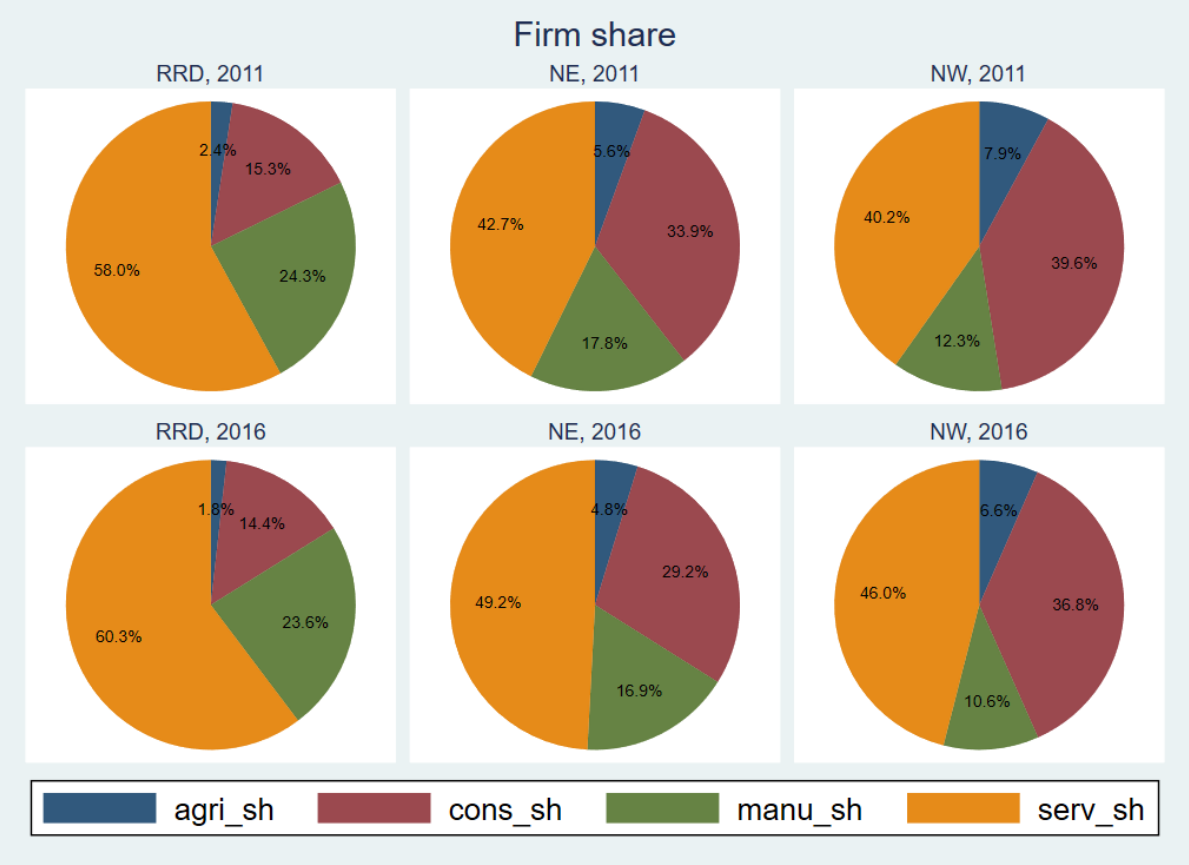
# Agglomeration pattern: Krugman INDEX

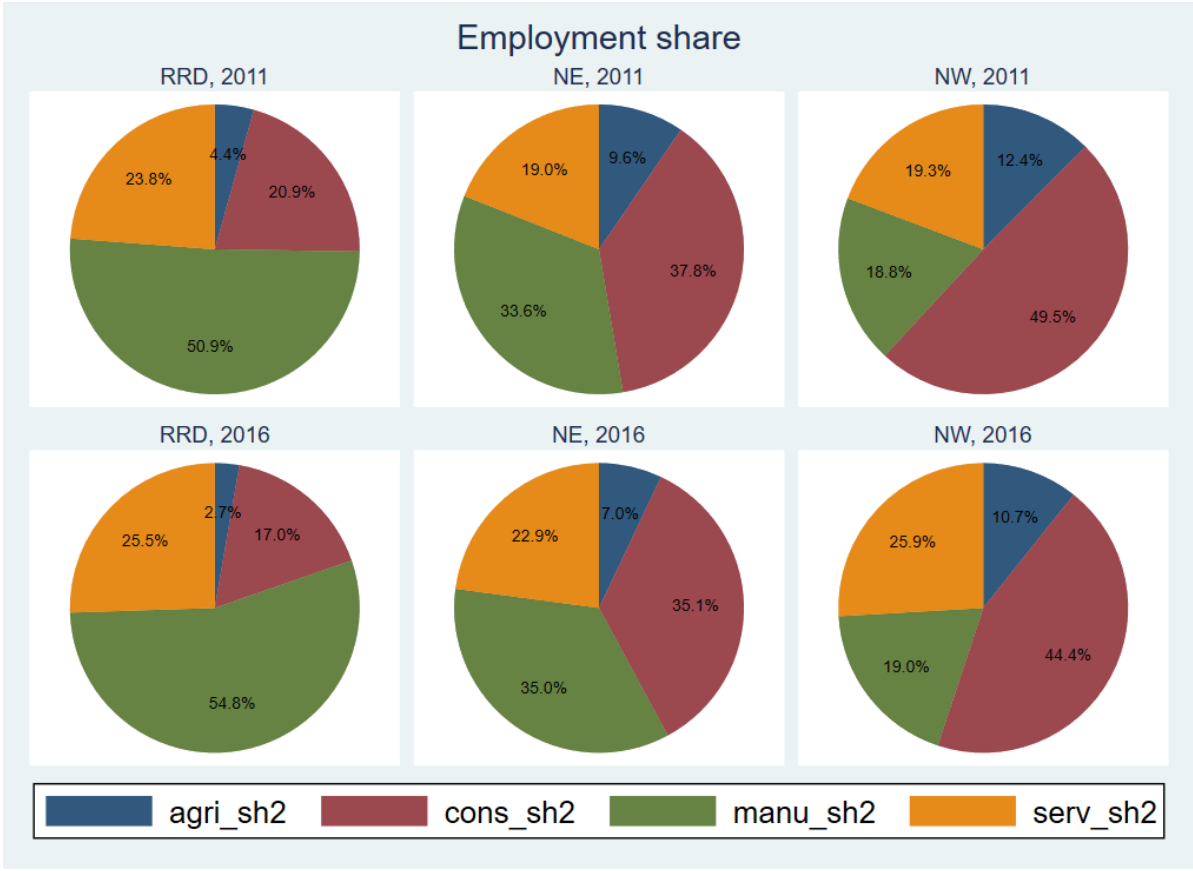
- Mountainous districts tend to have higher value of Krugman index due to relatively higher employment share in agricultural sector
- The Krugman index in these districts have changed dramatically during the period 2011-2016
- But no significant difference in the change between near-GMS districts and others.




Note: Krugman index is estimated using total district employment from the VES (registered firms)

# Agglomeration pattern: Firm and Employment Share




Note: Data is at the district level. Shares of firm and employment are calculated using VES data. Numbers shown in pie charts are average shares within each region.

# Baseline Results: Agglomeration of Firms

| | Log (number of firms) | | | | | | |
|---|---|---|---|---|---|---|---|
| VARIABLES | All | All | Manufacturing | Service | Small | Medium | Large |
| *A. RRD region* | | | | | | | |
| Log MPI (t-1) | 0.325*** | 0.141*** | 0.108*** | 0.129*** | 0.184*** | 0.023 | 0.036 |
| | (0.031) | (0.036) | (0.041) | (0.040) | (0.045) | (0.044) | (0.032) |
| Emp. Share in industrial zone (t-1) | 0.061 | 0.028 | 0.115*** | 0.005 | 0.048 | 0.029 | 0.149*** |
| | (0.050) | (0.029) | (0.042) | (0.032) | (0.033) | (0.045) | (0.045) |
| Log distance to GMS highway | -0.072** | | | | | | |
| | (0.030) | | | | | | |
| Observations | 13,245 | 13,208 | 13,208 | 13,208 | 13,208 | 13,208 | 13,208 |
| R-squared | 0.765 | 0.980 | 0.956 | 0.974 | 0.970 | 0.948 | 0.928 |
| Number of communes | 2,329 | 2,329 | 2,329 | 2,329 | 2,329 | 2,329 | 2,329 |
| *B. NE & NW regions* | | | | | | | |
| Log MPI (t-1) | 0.106*** | 0.046 | -0.054 | 0.003 | 0.077** | -0.400*** | 0.075 |
| | (0.029) | (0.043) | (0.046) | (0.037) | (0.031) | (0.080) | (0.071) |
| Emp. Share in industrial zone (t-1) | -0.092 | 0.159*** | 0.185*** | 0.150*** | 0.300*** | 0.294*** | -0.013 |
| | (0.096) | (0.051) | (0.050) | (0.050) | (0.056) | (0.097) | (0.083) |
| Log distance to GMS highway | -0.153*** | | | | | | |
| | (0.039) | | | | | | |
| Observations | 6,249 | 6,185 | 6,185 | 6,185 | 6,185 | 6,185 | 6,185 |
| R-squared | 0.487 | 0.954 | 0.947 | 0.913 | 0.888 | 0.766 | 0.735 |
| Number of communes | 1,225 | 1,225 | 1,225 | 1,225 | 1,225 | 1,225 | 1,225 |
| *Note: fixed effects* | | | | | | | |
| District FE & time trend | YES | NO | NO | NO | NO | NO | NO |
| Commune FE | NO | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES | YES |

***p<0.01, **p<0.05, *p<0.1. Standard errors, clustered at the level of fixed effect, are reported in the bracket.

# First-difference Model: First-stage Results

| | Log (MPI) | |
|---|---|---|
| VARIABLES | RRD | NE & NW |
| Land elevation in 1999 x log (Distance to GMS) | -0.022 | -0.019*** |
| | (0.014) | (0.005) |
| Avg. temperature in 1999 x log (Distance to GMS) | 0.087*** | |
| | (0.026) | |
| Avg. sunshine in 1999 x log (Distance to GMS) | -0.083*** | |
| | (0.016) | |
| Avg. precipitation in 1999 x log (Distance to GMS) | -0.025*** | |
| | (0.006) | |
| Number of U.S. bombs x log (Distance to GMS) | | 0.019*** |
| | | (0.004) |
| Log (Distance to GMS highway) | -0.010 | 0.044*** |
| | (0.007) | (0.016) |
| Observations | 2,194 | 974 |
| R-squared | 0.482 | 0.373 |
| Region FE | YES | YES |

***p<0.01, **p<0.05, *p<0.1. Standard errors, clustered at the district level, are reported in the bracket. All instruments except the distance to GMS highway are in z-score.

# First-difference Results: Agglomeration of Firms

| | ΔLog (number of firms) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| VARIABLES | All | All | IV All | IV Manufacturing | IV Service | IV Small | IV Medium | IV Large |
| *A. RRD region* | | | | | | | | |
| ΔLog MPI | 0.335 | 0.325 | 0.845** | 0.806** | 0.738* | 0.555 | 0.600** | 0.494*** |
| | (0.264) | (0.267) | (0.412) | (0.341) | (0.385) | (0.434) | (0.304) | (0.124) |
| Log employment in 2011 | | -0.008 | -0.006 | 0.003 | 0.021** | 0.029*** | -0.036*** | 0.001 |
| | | (0.008) | (0.008) | (0.008) | (0.009) | (0.009) | (0.008) | (0.005) |
| Observations | 2,194 | 2,194 | 2,194 | 2,194 | 2,194 | 2,194 | 2,194 | 2,194 |
| R-squared | 0.012 | 0.013 | 0.005 | 0.026 | 0.011 | 0.019 | 0.018 | 0.010 |
| 1st stage test statistics | | | | | | | | |
| Kleibergen-Paap F statistic | ... | ... | 35.72 | 35.72 | 35.72 | 35.72 | 35.72 | 35.72 |
| Hansen's J-statistics (p-value) | ... | ... | 0.00696 | 0.00660 | 0.0661 | 0.0608 | 0.00885 | 0.0364 |
| *B. NE & NW regions* | | | | | | | | |
| ΔLog MPI | 0.297** | 0.294** | 0.721** | 0.827** | 0.240 | 0.662** | 0.334 | 0.146 |
| | (0.132) | (0.120) | (0.286) | (0.324) | (0.254) | (0.329) | (0.250) | (0.104) |
| Log employment in 2011 | | 0.041*** | 0.037*** | 0.007 | 0.055*** | 0.088*** | -0.049*** | -0.002 |
| | | (0.009) | (0.010) | (0.009) | (0.010) | (0.012) | (0.009) | (0.008) |
| Observations | 1,042 | 1,042 | 974 | 974 | 974 | 974 | 974 | 974 |
| R-squared | 0.027 | 0.046 | 0.019 | -0.004 | 0.034 | 0.067 | 0.042 | 0.026 |
| 1st stage test statistics | | | | | | | | |
| Kleibergen-Paap F statistic | ... | ... | 12.64 | 12.64 | 12.64 | 12.64 | 12.64 | 12.64 |
| Hansen's J-statistics (p-value) | ... | ... | 0.552 | 0.456 | 0.554 | 0.193 | 0.134 | 0.00108 |
| *Note: fixed effects* | | | | | | | | |
| Region FE | YES | YES | YES | YES | YES | YES | YES | YES |
| With basic controls | YES | YES | YES | YES | YES | YES | YES | YES |

***p<0.01, **p<0.05, *p<0.1. Standard errors, clustered at the district level, are reported in the bracket.

# First-difference results: Employment Share

| | ΔShare of sector employments | |
|---|---|---|
| | IV | IV |
| VARIABLES | Manufacturing | Service |
| *A. RRD region* | | |
| ΔLog MPI | 0.100 | -0.107 |
| | (0.090) | (0.107) |
| Emp. share in industrial zone in 2011 | 0.057* | -0.077*** |
| | (0.031) | (0.020) |
| Log employment in 2011 | -0.015*** | 0.030*** |
| | (0.004) | (0.005) |
| Constant | 0.208*** | -0.160*** |
| | (0.028) | (0.033) |
| Observations | 2,159 | 2,159 |
| R-squared | 0.055 | 0.034 |
| 1st stage test statistics | | |
| Kleibergen-Paap F statistic | 35.88 | 35.88 |
| Hansen's J-statistics (p-value) | 0.269 | 0.0899 |
| *B. NE & NW regions* | | |
| ΔLog MPI | 0.351** | -0.136 |
| | (0.174) | (0.147) |
| Emp. share in industrial zone in 2011 | 0.035 | -0.031 |
| | (0.072) | (0.048) |
| Log employment in 2011 | -0.018*** | 0.030*** |
| | (0.007) | (0.007) |
| Constant | 0.098** | -0.010 |
| | (0.042) | (0.046) |
| Observations | 899 | 899 |
| R-squared | -0.020 | 0.031 |
| 1st stage test statistics | | |
| Kleibergen-Paap F statistic | 11.30 | 11.30 |
| Hansen's J-statistics (p-value) | 0.263 | 0.727 |
| *Note: fixed effects* | | |
| Region FE | YES | YES |
| With basic controls | YES | YES |

***p<0.01, **p<0.05, *p<0.1. Standard errors, clustered at the district level, are reported in the bracket.

# Dividends on Firms (by Industry)

| | Log (real value-added per employee) | | | Profit margins | | |
|---|---|---|---|---|---|---|
| VARIABLES | Manufacturing | Service: transport & related | Service: retail & wholesale | Manufacturing | Service: transport & related | Service: retail & wholesale |
| *A. RRD region* | | | | | | |
| Log MPI (t-1) | 0.193* | -0.113 | -0.706*** | 0.078*** | 0.079** | 0.067*** |
| | (0.114) | (0.119) | (0.118) | (0.030) | (0.037) | (0.023) |
| Observations | 10,192 | 6,624 | 11,139 | 9,938 | 6,451 | 10,805 |
| R-squared | 0.660 | 0.563 | 0.696 | 0.346 | 0.326 | 0.384 |
| *B. NE region* | | | | | | |
| Log MPI (t-1) | -0.027 | -0.169 | 0.535** | 0.070 | 0.005 | 0.026 |
| | (0.192) | (0.246) | (0.211) | (0.051) | (0.061) | (0.029) |
| Observations | 2,427 | 1,389 | 2,588 | 2,386 | 1,365 | 2,555 |
| R-squared | 0.671 | 0.550 | 0.639 | 0.429 | 0.383 | 0.379 |
| *C. NW region* | | | | | | |
| Log MPI (t-1) | 0.432** | 0.465* | 0.507*** | 0.207*** | -0.027 | 0.044 |
| | (0.208) | (0.273) | (0.178) | (0.061) | (0.066) | (0.045) |
| Observations | 1,089 | 623 | 1,203 | 1,030 | 585 | 1,188 |
| R-squared | 0.653 | 0.516 | 0.561 | 0.363 | 0.343 | 0.333 |
| *Note: fixed effects* | | | | | | |
| Commune FE | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES |

***p<0.01, **p<0.05, *p<0.1. Standard errors, clustered at the level of fixed effect, are reported in the bracket.

# Dividends on Firms (by Firm Size)

| | Log (real value-added per employee) | | | Profit margins | | |
|---|---|---|---|---|---|---|
| VARIABLES | Small | Medium | Large | Small | Medium | Large |
| *A. RRD region* | | | | | | |
| Log MPI (t-1) | -0.425*** | -0.140 | 0.097 | 0.150*** | -0.007 | -0.023 |
| | (0.122) | (0.088) | (0.093) | (0.034) | (0.016) | (0.018) |
| Observations | 12,461 | 10,167 | 4,292 | 12,200 | 10,022 | 4,269 |
| R-squared | 0.669 | 0.685 | 0.753 | 0.381 | 0.355 | 0.476 |
| *B. NE region* | | | | | | |
| Log MPI (t-1) | 0.408** | 0.171 | -0.095 | 0.013 | -0.016 | 0.049 |
| | (0.182) | (0.164) | (0.337) | (0.039) | (0.025) | (0.043) |
| Observations | 3,452 | 2,292 | 791 | 3,373 | 2,276 | 797 |
| R-squared | 0.594 | 0.615 | 0.815 | 0.338 | 0.400 | 0.580 |
| *C. NW region* | | | | | | |
| Log MPI (t-1) | 0.394*** | 0.212 | 0.139 | 0.053 | -0.007 | 0.018 |
| | (0.152) | (0.143) | (0.246) | (0.037) | (0.039) | (0.033) |
| Observations | 1,782 | 1,359 | 350 | 1,652 | 1,335 | 330 |
| R-squared | 0.519 | 0.628 | 0.819 | 0.408 | 0.432 | 0.541 |
| *Note: fixed effects* | | | | | | |
| Commune FE | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES |

***p<0.01, **p<0.05, *p<0.1. Standard errors, clustered at the level of fixed effect, are reported in the bracket.

# Effect of Market Potential on Average Wage (by Region)

| | Log (real average wage) | | | | | | |
|---|---|---|---|---|---|---|---|
| VARIABLES | All | All | Manufacturing | Service | Small | Medium | Large |
| *A. RRD region* | | | | | | | |
| Log MPI (t-1) | -0.017 | -0.691*** | -0.683*** | -0.967*** | -1.253*** | -0.688*** | 0.021 |
| | (0.018) | (0.082) | (0.101) | (0.111) | (0.128) | (0.097) | (0.060) |
| Log distance to GMS highway | -0.032*** | | | | | | |
| | (0.012) | | | | | | |
| Observations | 13,059 | 13,014 | 10,223 | 11,895 | 12,517 | 10,168 | 4,330 |
| R-squared | 0.323 | 0.484 | 0.498 | 0.459 | 0.396 | 0.473 | 0.609 |
| *B. NE region* | | | | | | | |
| Log MPI (t-1) | 0.040** | 0.026 | 0.126 | -0.111 | -0.092 | -0.010 | 0.137 |
| | (0.020) | (0.105) | (0.156) | (0.105) | (0.112) | (0.127) | (0.221) |
| Log distance to GMS highway | -0.067*** | | | | | | |
| | (0.024) | | | | | | |
| Observations | 3,826 | 3,775 | 2,433 | 2,942 | 3,458 | 2,319 | 810 |
| R-squared | 0.267 | 0.573 | 0.597 | 0.578 | 0.527 | 0.539 | 0.669 |
| *C. NW region* | | | | | | | |
| Log MPI (t-1) | 0.010 | 0.172* | 0.090 | 0.055 | 0.194* | 0.158 | 0.515** |
| | (0.024) | (0.094) | (0.168) | (0.100) | (0.109) | (0.116) | (0.248) |
| Log distance to GMS highway | -0.070 | | | | | | |
| | (0.050) | | | | | | |
| Observations | 2,092 | 2,061 | 1,082 | 1,448 | 1,792 | 1,397 | 351 |
| R-squared | 0.232 | 0.560 | 0.562 | 0.566 | 0.478 | 0.568 | 0.666 |
| *Note: fixed effects* | | | | | | | |
| District FE & time trend | YES | NO | NO | NO | NO | NO | NO |
| Commune FE | NO | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES | YES |

***p<0.01, **p<0.05, *p<0.1. Standard errors, clustered at the level of fixed effect, are reported in the bracket.

# Difference-in-Difference (DID)

- Estimate direct & spillover effect of the GMS highway
- Define treatment by distance and transportation intensity (Li, Wu, Chen, 2017)
  - the impact will be pronounced among communes relying intensively on transport services
  - use local input-output table in 2007 to compute the average transport intensity in sector production
  - given each sector outputs, compute the weighted average transport intensity of each commune
- Treatment externality (spillovers to control communes) are estimated following Miguel and Kremer (2004) and Lu, Wang, and Zhu (2019)
- DID without spillover effect

$$y_{ct} = \lambda_c + \lambda_{dt} + \beta D_c \times POST + (X_c \times \lambda_t)' \gamma + \varepsilon_{ct}$$

  - where $D_c$ is one if commune $c$ is within 20km from the highway and have transportation intensity greater than the province's median/75th percentile value
  - $\lambda_c$, $\lambda_{dt}$ and $\lambda_t$ are commune, district-year, and year FEs. $X_c$ is commune characteristics (distance to Hanoi, total trade value, and employment share in industrial zone).

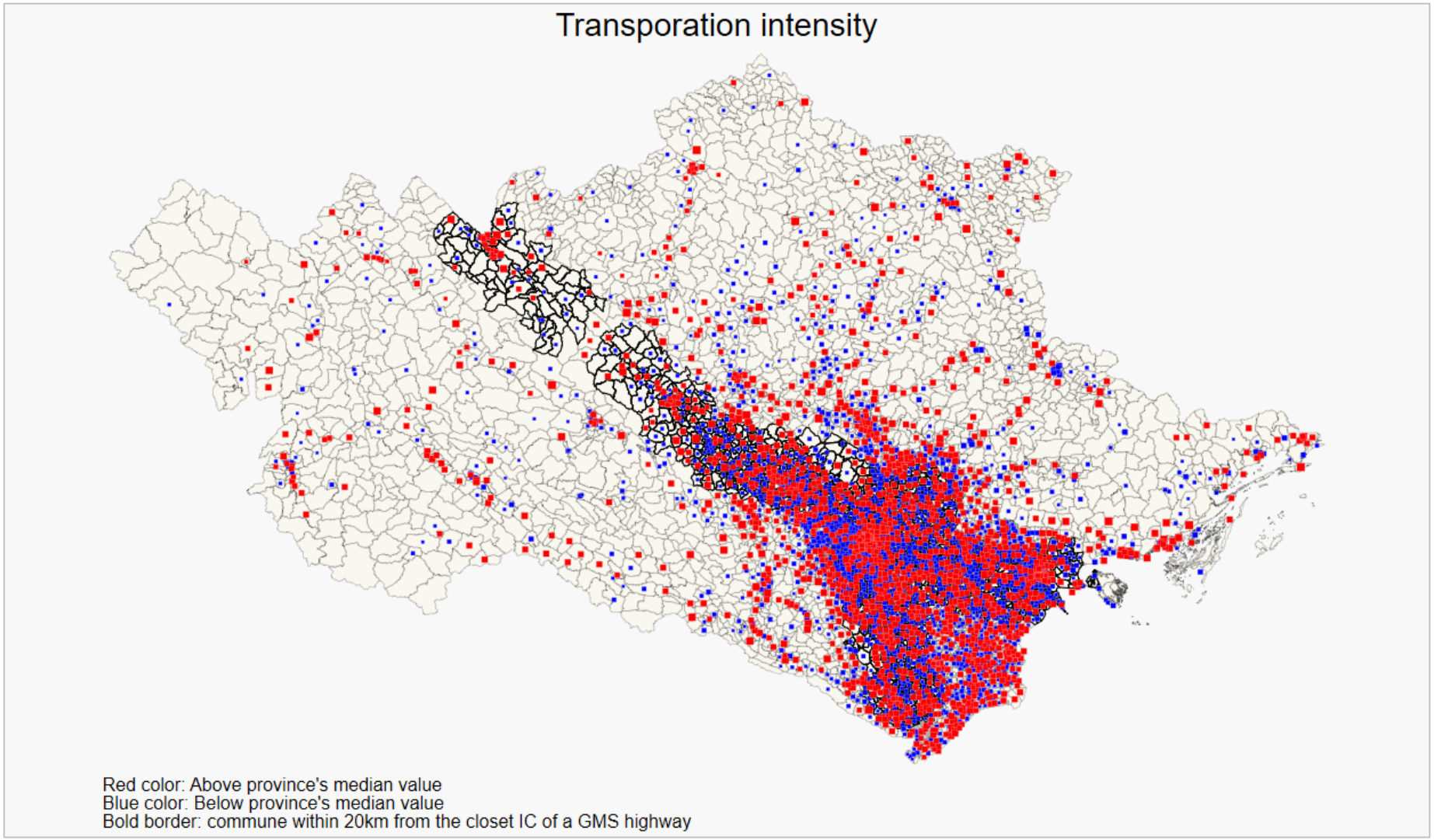
# Difference-in-Difference (DID)

- DID with spillover effect

$$y_{cdt} = \lambda_c + \lambda_t + \beta\left(D_{dt} \times D_c\right) + \sigma D_{dt} + \left(X_c \times \lambda_t\right)'\gamma + \varepsilon_{cdt}$$

  - where $D_{dt} = D_d \times POST$, $D_d$ is one if district $d$ has one or more treated communes.
  - $\beta$: additional direct effect on treated commune, $\sigma$: spillover effect on untreated communes.

# Transportation Intensity



# GMS's Agglomeration Effect

**A. Annual sample**

| **VARIABLES** | **Log (number of firms)** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Manufacturing | | | | Service: Transport & related | | | | Service: wholesale & trade | | | |
| | RRD region | | NE & NW region | | RRD region | | NE & NW region | | RRD region | | NE & NW region | |
| DcxPost | 0.035* | | 0.047 | | -0.055** | | 0.028 | | -0.051*** | | -0.124*** | |
| | (0.021) | | (0.033) | | (0.023) | | (0.044) | | (0.018) | | (0.042) | |
| Direct effect | | 0.048** | | 0.099*** | | -0.021 | | 0.068 | | -0.045** | | -0.075* |
| | | (0.021) | | (0.035) | | (0.026) | | (0.043) | | (0.021) | | (0.043) |
| Spillover effect | | 0.018 | | -0.048 | | 0.032 | | 0.009 | | -0.027 | | -0.011 |
| | | (0.021) | | (0.034) | | (0.025) | | (0.027) | | (0.027) | | (0.033) |
| Observations | 13,056 | 13,056 | 5,862 | 5,901 | 13,056 | 13,056 | 5,862 | 5,901 | 13,056 | 13,056 | 5,862 | 5,901 |
| R-squared | 0.961 | 0.956 | 0.917 | 0.902 | 0.959 | 0.953 | 0.926 | 0.913 | 0.971 | 0.967 | 0.947 | 0.934 |

**B. Cumulative over 5 year (2011-2016)**

| VARIABLES | Manufacturing | | | | Service: Transport & related | | | | Service: wholesale & trade | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RRD region | | NE & NW region | | RRD region | | NE & NW region | | RRD region | | NE & NW region | |
| DcxPost | 0.075** | | 0.128* | | -0.109*** | | -0.039 | | -0.048* | | -0.184*** | |
| | (0.034) | | (0.070) | | (0.033) | | (0.076) | | (0.027) | | (0.068) | |
| Direct effect | | 0.095*** | | 0.177*** | | -0.061 | | 0.011 | | -0.052* | | -0.118* |
| | | (0.032) | | (0.067) | | (0.037) | | (0.068) | | (0.031) | | (0.069) |
| Spillover effect | | 0.011 | | -0.035 | | 0.059 | | 0.049 | | -0.054 | | 0.040 |
| | | (0.034) | | (0.046) | | (0.037) | | (0.038) | | (0.038) | | (0.053) |
| Observations | 4,306 | 4,306 | 1,886 | 1,898 | 4,306 | 4,306 | 1,886 | 1,898 | 4,306 | 4,306 | 1,886 | 1,898 |
| R-squared | 0.960 | 0.954 | 0.920 | 0.904 | 0.959 | 0.951 | 0.929 | 0.920 | 0.972 | 0.967 | 0.952 | 0.942 |
| Commune FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| District x Year FE | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO |
| Year FE | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES |

***p<0.01, **p<0.05, *p<0.1. Standard errors in parantheses are clustered by district.

# GMS's Employment Effect

**A. Annual sample**

| **VARIABLES** | **Log employment** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Manufacturing | | | | Service: Transport & related | | | | Service: wholesale & trade | | | |
| | RRD region | | NE & NW region | | RRD region | | NE & NW region | | RRD region | | NE & NW region | |
| DcxPost | 0.136** | | 0.446*** | | -0.060 | | 0.054 | | -0.140*** | | -0.067 | |
| | (0.063) | | (0.104) | | (0.049) | | (0.110) | | (0.037) | | (0.099) | |
| Direct effect | | 0.145*** | | 0.419*** | | 0.002 | | 0.062 | | -0.142*** | | -0.018 |
| | | (0.054) | | (0.117) | | (0.044) | | (0.107) | | (0.034) | | (0.088) |
| Spillover effect | | 0.075 | | -0.186* | | 0.088 | | -0.008 | | 0.080 | | 0.006 |
| | | (0.061) | | (0.096) | | (0.054) | | (0.068) | | (0.052) | | (0.065) |
| Observations | 13,056 | 13,056 | 5,862 | 5,901 | 13,056 | 13,056 | 5,862 | 5,901 | 13,056 | 13,056 | 5,862 | 5,901 |
| R-squared | 0.918 | 0.913 | 0.905 | 0.894 | 0.928 | 0.921 | 0.900 | 0.887 | 0.938 | 0.932 | 0.907 | 0.892 |

**B. Cumulative over 5 year (2011-2016)**

| VARIABLES | Manufacturing | | | | Service: Transport & related | | | | Service: wholesale & trade | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RRD region | | NE & NW region | | RRD region | | NE & NW region | | RRD region | | NE & NW region | |
| DcxPost | 0.281*** | | 0.838*** | | -0.135* | | 0.000 | | -0.233*** | | -0.184 | |
| | (0.101) | | (0.179) | | (0.077) | | (0.184) | | (0.054) | | (0.166) | |
| Direct effect | | 0.292*** | | 0.732*** | | -0.047 | | 0.091 | | -0.255*** | | -0.082 |
| | | (0.085) | | (0.190) | | (0.075) | | (0.164) | | (0.050) | | (0.140) |
| Spillover effect | | 0.113 | | -0.247* | | 0.135 | | -0.052 | | 0.109 | | 0.155 |
| | | (0.093) | | (0.140) | | (0.088) | | (0.099) | | (0.073) | | (0.100) |
| Observations | 4,306 | 4,306 | 1,886 | 1,898 | 4,306 | 4,306 | 1,886 | 1,898 | 4,306 | 4,306 | 1,886 | 1,898 |
| R-squared | 0.915 | 0.909 | 0.898 | 0.885 | 0.929 | 0.921 | 0.901 | 0.890 | 0.943 | 0.937 | 0.916 | 0.905 |
| Commune FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| District x Year FE | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO |
| Year FE | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES |

***p<0.01, **p<0.05, *p<0.1. Standard errors in parantheses are clustered by district.

# GMS's Effect on Employment Share

**A. Annual sample**

| **VARIABLES** | **Share of sector employments** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Manufacturing | | | | Service: Transport & related | | | | Service: wholesale & trade | | | |
| | RRD region | | NE & NW region | | RRD region | | NE & NW region | | RRD region | | NE & NW region | |
| DcxPost | 0.039*** | | 0.079*** | | -0.023*** | | 0.005 | | -0.052*** | | -0.051* | |
| | (0.011) | | (0.022) | | (0.006) | | (0.014) | | (0.009) | | (0.028) | |
| Direct effect | | 0.037*** | | 0.065*** | | -0.015*** | | 0.003 | | -0.053*** | | -0.026 |
| | | (0.010) | | (0.022) | | (0.005) | | (0.011) | | (0.008) | | (0.026) |
| Spillover effect | | -0.010 | | -0.023 | | 0.009 | | 0.003 | | 0.004 | | 0.001 |
| | | (0.010) | | (0.018) | | (0.007) | | (0.006) | | (0.011) | | (0.018) |
| Observations | 13,055 | 13,055 | 5,857 | 5,896 | 13,055 | 13,055 | 5,857 | 5,896 | 13,055 | 13,055 | 5,857 | 5,896 |
| R-squared | 0.837 | 0.827 | 0.838 | 0.822 | 0.756 | 0.740 | 0.685 | 0.654 | 0.740 | 0.724 | 0.746 | 0.716 |

**B. Cumulative over 5 year (2011-2016)**

| VARIABLES | Manufacturing | | | | Service: Transport & related | | | | Service: wholesale & trade | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RRD region | | NE & NW region | | RRD region | | NE & NW region | | RRD region | | NE & NW region | |
| DcxPost | 0.080*** | | 0.178*** | | -0.042*** | | -0.013 | | -0.101*** | | -0.101** | |
| | (0.017) | | (0.037) | | (0.010) | | (0.021) | | (0.014) | | (0.043) | |
| Direct effect | | 0.081*** | | 0.143*** | | -0.029*** | | 0.000 | | -0.100*** | | -0.069* |
| | | (0.015) | | (0.038) | | (0.009) | | (0.017) | | (0.013) | | (0.040) |
| Spillover effect | | -0.020 | | -0.052* | | 0.014 | | -0.009 | | 0.000 | | 0.010 |
| | | (0.016) | | (0.027) | | (0.010) | | (0.010) | | (0.014) | | (0.028) |
| Observations | 4,306 | 4,306 | 1,886 | 1,898 | 4,306 | 4,306 | 1,886 | 1,898 | 4,306 | 4,306 | 1,886 | 1,898 |
| R-squared | 0.842 | 0.832 | 0.838 | 0.819 | 0.765 | 0.746 | 0.686 | 0.659 | 0.767 | 0.755 | 0.752 | 0.721 |
| Commune FE | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| District x Year FE | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO |
| Year FE | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES | NO | YES |

***p<0.01, **p<0.05, *p<0.1. Standard errors in parantheses are clustered by district.

# Validation with the NTL (First-difference & DID results)

| VARIABLES | ΔNighttime light | | | | Nighttime light, 2012-2016 | | | | Nighttime light, 2012-2019 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RRD region | | NE & NW region | | RRD region | | NE & NW region | | RRD region | | NE & NW region | |
| | OLS | IV | OLS | IV | | | | | | | | |
| Δlog MPI | 0.043 | 0.338 | 0.080 | 0.127 | | | | | | | | |
| | (0.122) | (0.221) | (0.059) | (0.136) | | | | | | | | |
| DcxPost | | | | | 0.015 | | 0.317 | | 0.206* | | 0.270 | |
| | | | | | (0.115) | | (0.290) | | (0.113) | | (0.194) | |
| Direct effect | | | | | | 0.663*** | | 0.654 | | 0.972*** | | 0.716** |
| | | | | | | (0.185) | | (0.439) | | (0.223) | | (0.345) |
| Spillover effect | | | | | | -1.013*** | | 0.076 | | -0.922* | | 0.224 |
| | | | | | | (0.376) | | (0.161) | | (0.469) | | (0.167) |
| Observations | 2,194 | 2,194 | 1,042 | 974 | 4,430 | 4,430 | 2,096 | 2,106 | 4,430 | 4,430 | 2,096 | 2,106 |
| R-squared | 0.204 | 0.194 | 0.303 | 0.299 | 0.986 | 0.968 | 0.943 | 0.903 | 0.981 | 0.962 | 0.947 | 0.910 |
| *1st stage statistics* | | | | | | | | | | | | |
| Kleibergen-Paap F statistic | | 11.87 | | 12.64 | | | | | | | | |
| Hansen's J-statistics (p-value) | | 1.35e-05 | | 0.0460 | | | | | | | | |
| *Note:* | | | | | | | | | | | | |
| Region FE | YES | YES | YES | YES | | | | | | | | |
| Commune FE | | | | | YES | YES | YES | YES | YES | YES | YES | YES |
| District x Year FE | | | | | YES | NO | YES | NO | YES | NO | YES | NO |
| Year FE | | | | | NO | YES | NO | YES | NO | YES | NO | YES |

Note: p<0.01, **p<0.05, *p<0.1. Same control variables are included as in the previous first-difference and DID specifications.