Supplementary Information
Cumulative environmental and employment impacts of the shale gas boom

Erin N. Mayfield, Jared L. Cohon, Nicholas Z. Muller, Inês M. L. Azevedo, Allen L. Robinson

Table of Contents
Supplementary Methods.................................................................2
Scope.................................................................................................2
Natural gas activity........................................................................3
Air quality model...........................................................................7
Climate change model...................................................................24
Employment model.......................................................................38
Supplementary References...........................................................50
Supplementary Methods

Scope

Figure 1. Map of Appalachian basin. The red outline is the extent of the Appalachian basin, as delineated by the U.S. Geological Survey.¹ The shaded blue region is the Marcellus play, the yellow region is the Utica play, and the green region is the intersection of the Marcellus and Utica plays.²,³

Figure 2. Shale gas withdrawals⁴-⁷, conventional natural gas withdrawals, and export and import volumes⁸ for Appalachia. Import volumes include net interstate receipts to the tristate, and export volumes include net deliveries to outside of the tristate. We exclude import and export volumes for transactions between Pennsylvania, Ohio, and West Virginia.

²,³
**Natural gas activity**

The following gives additional information regarding natural gas activity data sources, data cleaning, and treatment of missing or misreported data.

**Production**

For Pennsylvania, we include natural gas wells designated by the PA DEP as unconventional, which are horizontally or vertically drilled in unconventional formations and require stimulation through hydraulic fracturing. We exclude conventional wells, which produce from conventional formations and are vertically drilled; although conventional wells typically require hydraulic fracturing, they do not require the volume of fluids typically required for unconventional wells. Due to changing reporting requirements, annual production for 2010 cannot be extracted from the reported datasets; thus, we simulate 2010 production by interpolating production per well between 2009 and 2011, and account for the year in which the well was spud. For West Virginia, we include all wells designated in well permits by the WV DEP as horizontally configured and/or as having the Marcellus or Utica formations as the target producing formation. For Ohio, we include all wells designated by the ODNR as horizontal shale and/or Utica wells.

Figure 3 depicts the cumulative production aggregated by county, and Figure 4 depicts the time series of production aggregated by county. Most unconventional natural gas wells drilled in Pennsylvania between 2004 and 2016 produce from the Marcellus shale formation, with production from the Utica shale beginning in 2012 and increasingly thereafter, although many wells produce from or penetrate multiple formations. Production has continued to rise from 2004 to 2016, as shown in Figure 5, and there is high correspondence between the cleaned well-level production data and state-level estimates reported by the Energy Information Agency (EIA).

![Figure 3. Cumulative production from 2004 to 2016 for each shale gas well. The brown shades indicate the cumulative production aggregated at the county level.](image)
Figure 4. Maps of annual production from 2004 to 2016. The brown shades indicate the production aggregated by county.

Figure 5. Annual shale gas production from 2004 to 2016 for each state. Bars represent cleaned, well-level production data aggregated by state. The solid yellow line depicts shale production aggregated across Pennsylvania, Ohio, and West Virginia and reported at the state level by the EIA.

**Producing wells**

We derive spatially-resolved producing well. We include all wells that reported production in a given year. Producing well counts have continued to rise from 2004 to 2016, as shown in Figure 6.
Spud wells

We derive spatially-resolved spud well counts. The number of wells spud in a given year is estimated based on the reported or simulated year in which drilling commences. PA DEP and ODNR publish the spud date for most wells, while WV DEP does not. For wells without a reported spud year, we assume that the spud year is the year prior to when production is first reported.

Figure 7 depicts the annual number of spud wells. The first well was spud in Pennsylvania in 2004, while the first reported spud wells in West Virginia and Ohio were a few years later. The aggregate number of spud wells peaked in 2013, after initial rapid purchases and development of leases.
**Midstream and end use volumes**

Fuel consumption volumes from midstream and end use activity are depicted in Figure 8. Attributing air quality and climate change impacts from midstream and end use segments to shale production is challenging. For example, the volume of natural gas that enters a processing plant is indistinguishably derived from conventional or unconventional production sources. Moreover, flows of natural gas between segments and across state boundaries are inconsistent. As a proxy, we prorate reported natural gas volumes (or emissions), based on the percentage of shale gas out of total natural gas production, as provided in Figure 9.

![Figure 8](image1.png)

Figure 8. Midstream and end use natural gas volumes from 2004 to 2016 for Pennsylvania, Ohio, and West Virginia. Processed gas, lease fuel consumption, plant fuel consumption, pipeline and distribution consumption, and electric power, industrial, commercial, residential, and vehicle consumption within Pennsylvania, Ohio, and West Virginia.

![Figure 9](image2.png)

Figure 9. Prorate factor of shale to total natural gas production.
**Air quality model**

**Emissions model**

In the following sections, we describe emissions model formulations and assumptions. Table 1 is a summary of emissions modeling input parameters, including definitions, values, and data sources.

| Parameter                  | Parameter Definition                                                                 | Parameter Value | Units   | Source |
|----------------------------|--------------------------------------------------------------------------------------|-----------------|---------|--------|
| $CF_{completion}$          | emission control factor                                                             | VOC: 95         | %       | 18     |
| $CF_{condensate}$          | emission control factor                                                             | VOC: 95         | %       | 19     |
| $CF_{drilling,i}$          | emission control factor a                                                           | NOx: Triangular (10.95,30) | %     | 20.21  |
|                           |                                                                                     | PM2.5: Triangular (60.81,97) |       |        |
|                           |                                                                                     | VOC: Triangular (60.81, 97) |       |        |
| $CF_{fracking,i}$          | emission control factor a                                                           | NOx: Triangular (10.95,30) | %     | 20.21  |
|                           |                                                                                     | PM2.5: Triangular (60.81,97) |       |        |
|                           |                                                                                     | VOC: Triangular (60.81, 97) |       |        |
| $CF_{wellheadcompressor,i}$| emission control factor a                                                           | NOx: Triangular (15.50,95) | %     | 20     |
|                           |                                                                                     | VOC: Triangular (30.60, 95) |       |        |
| $DT_{drilling,t}$          | time to drill a single well                                                         | Triangular (14,30,35) | days    | 20     |
| $DPW_{fracking}$           | distance between public water source and well site                                  | Uniform (4.9, 27.3) | mi      | 22     |
| $DSW_{fracking}$           | distance between surface water source and well site                                 | Uniform (0, 9.9) | mi      | 22     |
| $DWW_{fracking,t}$         | distance between well site and wastewater site                                      | Time-variant (see Figure 11) | mi     | 23     |
| $EF_{completion,uncontrolled}$| VOC emissions factor for well completion                                           | Empirical (see Figure 10) | metric ton/well | 20     |
| $ET_{drilling}$            | percentage of the time the drilling equipment operates                              | Triangular (20,50,100) | %       | 20.24  |
| $F_{condensate}$           | VOC emissions per volume of condensate production without emission controls         | Empirical (see Figure 10) | g/bbl  | 25     |
| $F_{drilling,i}$           | emissions factor for all drilling rig engines (in base year of 2009)                | Empirical (see Figure 10) | g/hp-hr | 20     |
| $F_{fracking,i}$           | emissions factor for pump engine for fracturing                                     | Empirical (see Figure 10) | g/hp-hr | 20     |
| $F_{fugitives,c}$          | fugitive TOC emission factor for each component type c                              | valves: 4.5     | g/hr    | 20     |
|                           |                                                                                     | connectors: 0.2 |         |        |
|                           |                                                                                     | OEL: 2         |         |        |
|                           |                                                                                     | flanges: 0.39  |         |        |
| $F_{heater,i}$             | emissions factor per heater                                                         | NOX: Triangular (0.022,0.045,0.091) | g/scf | 20     |
|                           |                                                                                     | VOC:Triangular (0.0013,0.005,0.003) |       |        |
| $F_{fracking,i}$           | emission factor per mile of heavy duty diesel trucks                                | Empirical (see Figure 10) | g/mi   | 20     |
| $F_{pneumatics}$           | VOC emissions per pneumatic device per year                                         | Empirical (see Figure 10) | g/device/yr | 26     |
| $F_{wellheadcompressor,i}$ | emissions factor for wellhead compressor engines                                   | Empirical (see Figure 10) | g/hp-hr | 20     |
| $FC_{completion,t}$         | fraction of well completions in which emission controls are used                    | Time-variant points | -/-    | 27,28  |
| $FC_{wellheadcompressor,t}$ | fraction of wells with compressors                                                  | 3                | %       |        |
| $FT_{drilling,t}$          | cumulative fleet turnover fraction (relative to base year of 2009)c                 | Time-variant points (see Figure 11) | %     | 29-31  |
| $FT_{fracking,t}$          | cumulative fleet turnover fraction (relative to base year of 2009)d                 | Time-variant points (see Figure 11) | %     | 29-31  |
| $HHV_{heater}$             | heating value of natural gas                                                        | 1000            | BTU/cf  |        |
| $HP_{compressorstation}$   | horsepower per unit of production                                                   | Uniform(0.125,0.15) | hp/BCF | 20     |
| $HP_{drilling}$            | total horsepower of all engines on the drilling rig                                | Triangular (2000,7000,4260) | hp     | 20     |
| $HP_{fracking}$            | total horsepower of pump engine for fracturing                                      | Triangular (35,000, 40,000, 45,000) | hp-hr | 20     |
| $HP_{wellheadcompressor}$  | total horsepower of wellhead compressor engine                                      | Triangular (30, 101, 242) | hp-hr  | 20.24  |
| $LF_{compressorstation}$   | average load factor of compressor station engine                                   | Uniform(40,80) | %       | 20     |
| $LF_{drilling}$            | average load factor for all engines on drilling rig                                | Triangular (26,56,90) | %       | 20.32  |
| $LF_{fracking}$            | average load factor of pump engine for fracturing                                  | 0.5              | %       | 20.33  |
| $LF_{wellheadcompressor}$  | average load factor of wellhead compressor engine                                  | Empirical       | %       | 20     |
Table 1 (continued). Emissions modeling parameter values, definitions, and data sources.

| Parameter | Parameter Definition | Parameter Value | Units | Source |
|-----------|----------------------|-----------------|-------|--------|
| $N_{fugitives,c}$ | number of components of type c per well | valves: 15 connectors: 43 OEL: 5 flanges: 25 | componen| $^{24,34}$ |
| $N_{heater}$ | number of heaters per well | Triangular (0.0, 63.1, 1.1) | heaters/well | $^{24}$ |
| $N_{pneumatics}$ | total number of pneumatic device per well | Empirical | devices/well | $^{26}$ |
| $P_{condensate,site,s,j,t}$ | condensate production per well per year | Empirical (see Figure 10) | bbl/well/yr | |
| $PW_{trucking}$ | percentage of water sourced (excluding reused/recycled water) from public water supplies | 0.2 | % | $^{22}$ |
| $PSW_{trucking}$ | percentage of water sourced (excluding reused/recycled water) from surface water supplies | 0.8 | % | $^{22}$ |
| $Q_{heater}$ | heater throughput | 106 | BTU/hr | $^{20}$ |
| $R_{trucking,t}$ | rate of reuse/recycling of wastewater | Time-variant points (see Figure 11) | % | $^{23}$ |
| $RE_{commercial,state,i,j,t}$ | reported (or simulated) county level emissions from commercial end use | Time-varying | metric tons | $^{35}$ |
| $RE_{distribution,state,i,j,t}$ | reported (or simulated) emissions from distribution compressor stations | Time-varying | metric tons | $^{35}$ |
| $RE_{electric,state,i,j,t}$ | reported (or simulated) emissions from electric utility facilities | Time-varying | metric tons | $^{35-37}$ |
| $RE_{industrial,state,i,j,t}$ | reported (or simulated) county level emissions from industrial end use | Time-varying | metric tons | $^{35}$ |
| $RE_{processing,state,i,j,t}$ | reported (or simulated) emissions from processing facilities | Time-varying | metric tons | $^{35}$ |
| $RE_{residential,state,i,j,t}$ | reported (or simulated) emissions from residential end use | Time-varying | metric tons | $^{35}$ |
| $RE_{transmission,state,i,j,t}$ | reported (or simulated) emissions from transmission compressor stations | Time-varying | metric tons | $^{35}$ |
| $S_{fracking,t}$ | number stages to fracture well | Triangular (4, 15, 33) | stages | $^{20}$ |
| $T_{compressorstation}$ | number of hours engine operates per day | 24 | hrs | $^{20}$ |
| $T_{fugitives}$ | annual operating hours per well | 8760 | hrs/year | $^{34}$ |
| $T_{heater}$ | number of hours heater operates per year | Triangular (126, 2982, 4601) | hrs/yr | $^{24}$ |
| $T_{wellhead,compressor}$ | number of hours engine operates per year | 8760 | hrs/yr | $^{20}$ |
| $U_{state,t}$ | fraction of total unconventional production out of total production | Time-varying (ranges from 0 to 99%) | % | $^{4-7}$ |
| $TC_{trucking}$ | truck fluid capacity | 4620 | gal | $^{38}$ |
| $UW_{trucking}$ | water usage per well | Triangular (1.00, 4.05, 6.97) | million gal/well | $^{22}$ |
| $UWW_{trucking}$ | wastewater produced per well | Triangular (1.22, 1.38, 1.64) | million gal/well | $^{39}$ |
| $VF_{fugitives}$ | VOC fraction of natural gas | 0.02 | - | $^{20}$ |
| $W_{producing,j,t}$ | number of spud wells | Time-varying (refer to section 2) | wells | $^{4-6}$ |
| $W_{spud,j,t}$ | number of spud wells | Time-varying (refer to section 2) | wells | $^{4-6}$ |

a We assume ignition timing retard and selective catalytic reduction for NOx, diesel particulate filters for PM2.5, diesel oxidation catalysts for VOCs.

b To develop an empirical VOC emission factor distribution, we use whole gas emission factors and mole fractions of propane, butane, and higher hydrocarbons for a subset of 53 wells in the Appalachian region, as reported in Allen et al. (2014). While the Allen et al. study measures emissions in multiple basins, we subset the reported data for those devices measured within the Appalachian region because of that finding that there is great variability in regional emissions resulting from differences in controller type (e.g., continuous and intermittent venting), frequency of actuation (e.g., more actuation in wet areas), and different applications (e.g., separators, process heaters). We convert reported emission factors in terms of scf per hour to grams per year, under standard conditions (14.7 psia and 70°F) and assuming the pneumatic devices are operating 8760 hours per year.

c We assume the following to derive the fleet turnover curve: a normally distributed scrappage curve, average load factor of 0.43 for diesel bore / drill rigs, annual activity of 466 hours per year for diesel bore / drill rigs, an average life of large diesel engines of 7000 hours, and a 3.7% growth rate for diesel industrial engines.

d We assume the following to derive the fleet turnover curve: a normally distributed scrappage curve, average load factor of 0.43 for diesel other oil equipment, annual activity of 1281 hours per year for other oil equipment, an average life of large diesel engines of 7000 hours, and a 3.7% growth rate for diesel industrial engines.
Figure 10. Empirical distributions used in air quality emissions modeling. Drilling emission factor distributions for (A) NO\textsubscript{x}, (B) VOC, and (C) PM\textsubscript{2.5}. Hydraulic fracturing emission factor distributions for (D) NO\textsubscript{x}, (E) VOC, and (F) PM\textsubscript{2.5}. Trucking emission factor distributions for (G) NO\textsubscript{x}, (H) VOC, and (I) PM\textsubscript{2.5}. Completion emission factor distribution for (J) VOC. Wellhead compressor emission factor distributions for (K) NO\textsubscript{x}, (L) VOC, and (M) PM\textsubscript{2.5}. Pneumatic devices emission factor distribution for (N) VOC. Blue lines and dots are empirical emission factor distributions (without emission controls). Red lines are emission factor distributions, accounting for emission controls. Grey dashed lines are regulatory standards, including Tier 1, 2, and 4 non-road (hp>750) diesel engine standards and NSPS for spark-ignition natural gas engines.
Preproduction

We estimate drilling emissions \( E_{\text{drilling},i,j,t} \) for each species \( i \), source location \( j \), and year \( t \), using a modification of the approach used in Bar-Ilan et al. (2008), as follows:

\[
E_{\text{drilling},i,j,t} = W_{\text{spud},j,t} \cdot \left[ FT_{\text{drilling},t} \cdot EF_{\text{drilling,controlled},i,t} + (1 - FT_{\text{drilling},t}) \cdot EF_{\text{drilling,uncontrolled},i,t} \right]
\]

(1)

where \( W_{\text{spud},j,t} \) is the number of spud wells and \( FT_{\text{drilling},t} \) is the fleet turnover fraction. \( EF_{\text{drilling,controlled},i,t} \) and \( EF_{\text{drilling,uncontrolled},i,t} \) are the emission factors per well for each year when emission controls are or are not employed, respectively, and are given by:

\[
EF_{\text{drilling,controlled},i,t} = F_{\text{drilling},i} \cdot (1 - CF_{\text{drilling},i}) \cdot HP_{\text{drilling}} \cdot LF_{\text{drilling}} \cdot DT_{\text{drilling},t} \cdot ET_{\text{drilling}}
\]

(2)

\[
EF_{\text{drilling,uncontrolled},i,t} = F_{\text{drilling},i} \cdot HP_{\text{drilling}} \cdot LF_{\text{drilling}} \cdot DT_{\text{drilling},t} \cdot ET_{\text{drilling}}
\]

(3)

where \( HP_{\text{drilling}} \) is the horsepower of drilling rig engines, \( LF_{\text{drilling}} \) is the load factor of the drilling rig engines, \( DT_{\text{drilling},t} \) is the drilling time, and \( ET_{\text{drilling}} \) is the engine-on-time percentage. We use empirical distributions of engine emissions factors \( F_{\text{drilling},i} \) for a base year of 2009, which were compiled in Roy et al. (2014). We assume a fleet turnover fraction \( FT_{\text{drilling},t} \) and apply control factors \( CF_{\text{drilling},i} \), which together reflect the phase-in of Tier 2 and Tier 4 standards for non-road diesel engines. We develop emission factors \( EF_{\text{drilling,controlled},i,t} \) and \( EF_{\text{drilling,uncontrolled},i,t} \) representing the uncertainty in the systems-level mean rather than between-unit variability; we use...
a Monte Carlo simulation approach and find the mean and the upper and lower 95% confidence interval.

We estimate hydraulic fracturing emissions \( (E_{\text{fracking},i,j,t}) \), using a modification of the approach used in Bar-Ilan et al. (2008), as follows:\(^{24}\)

\[
E_{\text{fracking},i,j,t} = W_{\text{spud},j,t} \cdot [FT_{\text{fracking},t} \cdot EF_{\text{fracking,controlled},i,t} + (1 - FT_{\text{fracking},t}) \cdot EF_{\text{fracking,uncontrolled},i,t}]
\]

where \( W_{\text{spud},j,t} \) is the number of spud wells and \( FT_{\text{fracking},t} \) is the pump engine fleet turnover fraction. \( EF_{\text{fracking,controlled},i,t} \) and \( EF_{\text{fracking,uncontrolled},i,t} \) are the emission factors per spud well for each year when emission controls are or are not employed, respectively, and are given by:

\[
EF_{\text{fracking,controlled},i,t} = F_{\text{fracking},i} \cdot (1 - CF_{\text{fracking},i}) \cdot HP_{\text{drilling}} \cdot LF_{\text{drilling}} \cdot S_{\text{fracking},t}
\]

\[
EF_{\text{fracking,uncontrolled},i,t} = F_{\text{fracking},i} \cdot HP_{\text{drilling}} \cdot LF_{\text{drilling}} \cdot S_{\text{fracking},t}
\]

where \( HP_{\text{drilling}} \) is the horsepower of drilling rig engines, \( LF_{\text{drilling}} \) is the load factor of the drilling rig engines, \( T_{\text{drilling},t} \) is the drilling time, and \( ET_{\text{drilling}} \) is the engine-on-time percentage. We use empirical distributions of pump engine emissions factors \( (F_{\text{fracking},i}) \) for a base year of 2009, which were compiled in Roy et al. (2014). We incorporate time variant elements, including emission regulations. We assume a fleet turnover fraction \( (FT_{\text{fracking},t}) \) and apply control factors \( (CF_{\text{fracking},i}) \), which together reflect the phase-in of Tier 2 and Tier 4 standards for non-road diesel engines. We develop emission factors \( (EF_{\text{fracking,controlled},i,t} \) and \( EF_{\text{fracking,uncontrolled},i,t}) \) representing the uncertainty in the systems-level mean rather than between-unit variability; we use a Monte Carlo simulation approach and find the mean and the upper and lower 95% confidence interval.

Emissions from well completions are highly uncertain.\(^{40-42}\) We estimate VOC emissions from well completion \( (E_{\text{completion},i,j,t}) \) as follows:

\[
E_{\text{completion},i,j,t} = W_{\text{spud},j,t} \cdot [FC_{\text{completion},t} \cdot EF_{\text{completion,controlled}} + (1 - FC_{\text{completion},t}) \cdot EF_{\text{completion,uncontrolled},i,t}]
\]

where \( W_{\text{spud},j,t} \) is the number of spud wells and \( FC_{\text{completion},t} \) is the fraction wells with emission controls. VOC emission factors for wells with \( (EF_{\text{completion,controlled}}) \) and without emission controls \( (EF_{\text{completion,uncontrolled}}) \), where the controlled emission factor is given by:

\[
EF_{\text{completion,controlled}} = EF_{\text{completion,uncontrolled}} \cdot (1 - CF_{\text{completion}})
\]

We use an empirical distribution of emissions factors \( (EF_{\text{completion,uncontrolled},i,t}) \) for a base year of 2009, which were compiled in Roy et al. (2014). We attempt to reflect voluntary adoption of emission controls and rapidly evolving regulation, including the implementation of the NSPS subpart OOOO and National Emissions Standards for Hazardous Air Pollutants (NESHAP) standards, that require reduced emission completions (RECs) for hydraulically fractured wells. We assume a changing fraction of well completions with emission controls \( (FC_{\text{completion},t}) \); the penetration of emission controls over time is uncertain, although recent emission measurement studies suggest that most wells in Appalachia (for which measurements were taken) have emission controls; in a study conducted in 2012 by Allen et al. (2013), all five measured well completions in Appalachia had emission controls, and in a study conducted in 2014 by Omara et al. (2016), all four measured well completions had emission controls, with three of having installed RECs.\(^{27,28}\) The emission factors
represent the uncertainty in the systems-level mean rather than between-unit variability; we use a Monte Carlo simulation approach and find the mean and the upper and lower 95% confidence interval.

Emissions from trucking, including transport of drilling and fracturing water and wastewater \((E_{\text{trucking},i,t})\) are estimated as follows:

\[
E_{\text{trucking},i,t} = W_{\text{spud},i,t} \cdot EF_{\text{trucking},i,t}
\]

where \(W_{\text{spud},i,t}\) is the number of spud wells, and \(EF_{\text{trucking},i,t}\) is the emission factor, as given by:

\[
EF_{\text{trucking},i,t} = \left[ UWW_{\text{trucking}} \cdot DWW_{\text{trucking},t} + UW_{\text{trucking}} \cdot (DSW_{\text{trucking}} \cdot PSW_{\text{trucking}} + DPW_{\text{trucking}} \cdot PPW_{\text{trucking}}) \cdot (1 - R_{\text{trucking},t}) \right] / TC_{\text{trucking}} \cdot 2 \cdot F_{\text{trucking},i,t}
\]

where \(UWW_{\text{trucking}}\) and \(WU_{\text{trucking}}\) are the water and wastewater use per well, respectively. \(DWW_{\text{trucking},t}\), \(DSW_{\text{trucking}}\), and \(DPW_{\text{trucking}}\) are the distance per trip for transporting between the well site and the wastewater disposal, surface water source, and public water source, respectively. \(PSW_{\text{trucking}}\) and \(PPW_{\text{trucking}}\) are the percentage of water from surface and public water supplies, respectively. \(R_{\text{trucking},t}\) is the percentage of wastewater that is reused and \(TC_{\text{trucking}}\) is the capacity per truck. We use empirical distributions of emissions factors per mile for heavy duty diesel trucks \((F_{\text{trucking},i,t})\) compiled in Roy et al. (2014). We exclude trucking from other site operations, given that fluid transport is the most significant source of truck traffic. We incorporate time variant elements reflective of changing wastewater management practices; specifically, we reflect the declining distance traveled to wastewater disposal sites and the increasing portion of wastewater that is reused, resulting from changing regulation and increasing wastewater infrastructure. We assume that trucking is co-located within the county (or grid cell) that wells are located.

Production

Emissions from wellhead compressors \((E_{\text{wellheadcompressor},i,t})\) are estimated as follows:

\[
E_{\text{wellheadcompressor},i,t} = W_{\text{producing},i,t} \cdot FC_{\text{wellheadcompressor}} \cdot EF_{\text{wellheadcompressor},i,t}
\]

where \(W_{\text{producing},i,t}\) is the number of producing wells, \(FC_{\text{wellheadcompressor}}\) is the fraction of wells with compressors, and \(EF_{\text{wellheadcompressor},i,t}\) is the emissions factor per well, given by the following set of equations:

\[
EF_{\text{wellheadcompressor},i,t \geq 2011} = F_{\text{wellheadcompressor},i,t} \cdot HP_{\text{wellheadcompressor}} \cdot LF_{\text{wellheadcompressor}} \cdot T_{\text{wellheadcompressor}}
\]

\[
EF_{\text{wellheadcompressor},i,t < 2011} = (1 - CF_{\text{wellheadcompressor},i}) \cdot F_{\text{wellheadcompressor},i,t} \cdot HP_{\text{wellheadcompressor}} \cdot LF_{\text{wellheadcompressor}} \cdot T_{\text{wellheadcompressor}}
\]

where \(HP_{\text{wellheadcompressor}}\) is the total horsepower of wellhead compressor engines per well, \(LF_{\text{wellheadcompressor}}\) is the load factor of wellhead compressor engines, \(T_{\text{wellheadcompressor}}\) is the number of operating hours per year, and \(CF_{\text{wellheadcompressor},i}\) is the emission control factor. We modified the approach used in Roy et al. (2014) to incorporate time variant elements, including the fraction of wellheads that have compressors and emission regulations. We use empirical distributions of emissions factors for wellhead compressor engines \((F_{\text{wellheadcompressor},i,t})\) (for a base year of 2009) compiled in Roy et al. (2014). To estimate emissions in other years, we assume an increasing fraction of wellheads with compressor and apply control factors, reflecting the phase-in of NSPS subpart JJJJ Standards of Performance for Stationary Spark Ignition Internal Combustion
Engines for NO\textsubscript{x} and VOC. Although wellhead compressors are currently uncommon in the Appalachian Basin, production pressure declines as the field ages, necessitating the use of wellhead compressors. We assume all wellhead compressors after 2011 have emission controls, given that wellhead compressor engines have short lifespans and there is low penetration of wellhead compressors (<1%) prior to 2011.

Emissions from condensate volatilization ($E_{\text{condensate},j,t}$) from tanks are estimated as follows:

$$E_{\text{condensate},j,t} = \sum_{\text{site}} P_{\text{condensate,site},s,j,t} \cdot EF_{\text{condensate},s}$$  \hspace{1cm} (14)

where $P_{\text{condensate,site},s,j,t}$ is the well site-specific condensate production for a given spud year $s$, and $EF_{\text{condensate},s,t}$ is the emissions factor per unit condensate production, given by the following set of equations:

$$EF_{\text{condensate},s<2011} = F_{\text{condensate}}$$  \hspace{1cm} (15)

$$EF_{\text{condensate},s\geq2011} = F_{\text{condensate}} \cdot (1 - CF_{\text{condensate}})$$  \hspace{1cm} (16)

where $F_{\text{condensate}}$ is the VOC emission factor per unit condensate production without controls, and $CF_{\text{condensate}}$ is the emission control factor. We incorporate time variant elements, namely NSPS subpart OOOO, promulgated in 2012, for VOC emissions from storage vessels at production sites; the rule requires 95% control for all storage vessels emitting at least 6 tpy of VOC, which were constructed, modified, or reconstructed after August 2011. We use empirical distributions of emissions factors, representing uncontrolled condensate tank emissions compiled in Roy et al. (2014). We assume that all wells spud prior to 2011 are uncontrolled, and those spud in 2011 or later are controlled and comply with the NSPS 95% control factor.

Emissions from pneumatic devices ($E_{\text{pneumatics},i,j,t}$) are estimated as follows:

$$E_{\text{pneumatics},i,j,t} = W_{\text{producing},j,t} \cdot EF_{\text{pneumatics},i}$$  \hspace{1cm} (17)

where $W_{\text{producing},j,t}$ is the number of producing wells, and $EF_{\text{pneumatics},i}$ is the VOC emissions per well, as given by:

$$EF_{\text{pneumatics},i,j,t} = F_{\text{pneumatics},i} \cdot N_{\text{pneumatics}}$$  \hspace{1cm} (18)

We base the emission factor per device ($F_{\text{pneumatics},i}$) and number of devices per well ($N_{\text{pneumatics}}$) distributions on measurements from a study by Allen et al. (2014). The NSPS subpart OOOOa for VOC emissions from pneumatic controllers at production sites were promulgated in 2012; the rule sets a whole gas bleed rate limit of 6 scf/h for continuous bleed, natural gas-driven pneumatic controllers, which commenced construction after August 2011, and are located between the wellhead and the point at which the gas enters the transmission and storage segment. We do not account for any time-varying parameters, given field data for the Appalachian region reported in Allen et al. (2014) indicate most whole gas emission factors below the 6 scf/h standard and most pneumatic controllers are intermittent rather than continuous bleed; in addition, most wells in the Appalachian basin were constructed after the date put forth in the 2012 NSPS.

Heater emissions for each producing well ($E_{\text{heater},i,j,t}$) are estimated, using the approach described in Roy et al. (2014), as follows:

$$E_{\text{heater},i,j,t} = W_{\text{producing},j,t} \cdot EF_{\text{heater},i}$$  \hspace{1cm} (19)

where $W_{\text{producing},j,t}$ is the number of producing wells, and $EF_{\text{heater},i}$ is the emissions per well, as given by:
EF_{heater,i,j} = F_{heater,i} \cdot Q_{heater} \cdot \frac{T_{heater} \cdot N_{heater}}{HHV_{heater}} \tag{20}

where $Q_{heater}$ is the heater throughput, and $N_{pneumatics}$ is the number of pneumatic devices per well. We base the emission factor ($F_{pneumatics}$), $T_{heater}$ is the number operating hours per year, $HHV_{heater}$ is the higher heating value, and $N_{heater}$ is the number of heaters per well. We do not anticipate any time-varying parameters over the period of analysis.

Fugitive emissions ($E_{fugitives,i,t}$) associated with leaking valves, connectors, flanges, and open-ended lines (OEL), are estimated as follows:

$$E_{fugitives,i,t} = W_{producing,i,t} \cdot EF_{fugitives} \tag{21}$$

where $W_{producing,i,t}$ is the number of producing wells, and $EF_{fugitives}$ is the VOC emissions factor per well, as given by:

$$E_{fugitives,t} = T_{fugitives} \cdot VF_{fugitives} \cdot \sum_c (F_{fugitives,c} \cdot N_{fugitives,c}) \tag{22}$$

where $T_{fugitives}$ is the annual operating hours per well, $VF_{fugitives}$ is the VOC fraction, $F_{fugitives,c}$ is fugitive total organic carbon emission factor for each component type $c$, and $N_{fugitives,c}$ is the number of components of each type. We do not account for any time-varying parameters and assume point estimates for each input variable, given that this is a relatively very minor emission source. The NSPS subpart 0000a for fugitive VOC emissions from production were promulgated but will not take effect (depending upon ongoing regulatory actions) until after the time horizon of this analysis. In addition, Ohio EPA has regulated fugitive emissions from production operations since 2014, requiring quarterly detection at some sites.

**Processing**

Emissions from processing compressor stations ($E_{processing,i,j,t}$), including only those associated with natural gas liquids extraction facilities are estimated as follows:

$$E_{processing,i,j,t} = RE_{processing,\text{state},i,j,t} \cdot U_{\text{state},t} \tag{23}$$

where $RE_{processing,\text{state},i,j,t}$ is the reported (or simulated) emissions, and $U_{\text{state},t}$ is the proration factor based on the ratio of shale production to all production for each state. We use reported facility-level emissions and point source coordinate locations from the EPA National Emissions Inventory (NEI) for facilities with NAICS 4862 for years 2005, 2008, 2011, and 2014. For non-reporting years, we estimate emissions for each source location $j$ by linearly interpolating between years or extrapolating across years, using reported emissions aggregated to a given source location resolution (i.e., county or 36 x 36 km grid cell). While the NEI dataset is inclusive of compressor stations associated with natural gas liquids extraction facilities, it does not include other types of processing, including ethane crackers or wellhead processing such as glycol dehydration; these other sources are likely non-trivial at present and will be even more significant in the coming years.

**Transmission**

Emissions from transmission compressor stations ($E_{transmission,i,j,t}$) are estimated as follows:

$$E_{transmission,i,j,t} = RE_{transmission,\text{state},i,j,t} \cdot U_{\text{state},t} \tag{24}$$

where $RE_{transmission,\text{state},i,j,t}$ is the reported (or simulated) emissions. We use reported facility-level emissions and point source coordinate locations from the EPA NEI for facilities with NAICS 211112 for years 2005, 2008, 2011, and 2014. For non-reporting years, we estimate emissions for each source location $j$ by linearly interpolating between years or extrapolating across years, using
reported emissions aggregated to a given source location resolution (i.e., county or 36 x 36 km grid cell). While the NEI dataset is inclusive of compressor stations, it does not include pipelines.

**Distribution**

Emissions from distribution compressor stations \(E_{\text{distribution},i,j,t}\) are estimated as follows:

\[
E_{\text{distribution},i,j,t} = RE_{\text{distribution state},i,j,t} \cdot U_{\text{state},t}
\]  

(25)

where \(RE_{\text{distribution state},i,j,t}\) is the reported (or simulated) emissions. We use reported facility-level emissions and point source coordinate locations from the EPA NEI for facilities with NAICS 22121 for years 2005, 2008, 2011, and 2014. For non-reporting years, we estimate emissions for each source location \(j\) by linearly interpolating between years or extrapolating across years, using reported emissions aggregated to a given source location resolution (i.e., county or 36 x 36 km grid cell). While the NEI dataset is inclusive of compressor stations, it does not include pipelines.

**End use**

We estimate emissions from electric utilities \(E_{\text{electric},i,j,t}\) as follows:

\[
E_{\text{electric},i,j,t} = \sum_{\text{state}} RE_{\text{electric state},i,j,t} \cdot U_{\text{state},t}
\]  

(26)

where \(RE_{\text{electric state},i,j,t}\) is the reported (or simulated) emissions for each spatial unit \(j\).

A variety of sources report NO\(_x\) emissions from the electric power sector, each with varying sectoral and temporal coverage, as shown in Table 2. For NO\(_x\) emissions, we use reported plant-level emissions from the EPA Continuous Emissions Monitoring System (CEMS) because it provides estimates for all years from 2004 to 2016 and is reported at a high spatial resolution; we include facilities with a primary fuel type of pipeline natural gas and identified as electric utilities, including cogeneration. The CEMS dataset is inclusive of most electric power generation, as indicated by the relative heat input and generation across datasets (Table 3).

For VOC and PM\(_{2.5}\) emissions, we use reported facility-level emissions from NEI for years 2005, 2008, 2011, and 2014. To extract natural gas facility emissions data from the NEI dataset and coordinate locations from the eGRID dataset, we do a crosswalk of the NEI, eGRID, and CEMS datasets. For non-reporting years, we estimate emissions for each source location \(j\) by linearly interpolating between years or extrapolating across years, using reported emissions aggregated to a given source location resolution (i.e., county or 36 x 36 km grid cell).

### Table 2. Comparison of electric power emissions sector data sources.

| Reporting system | Sector Coverage | Subsector Coverage | Years | Spatial Resolution |
|------------------|-----------------|--------------------|-------|-------------------|
| CEMS             | Electric Power Industry (generators >25 MW) | Electric Utility; Industrial Boiler; Pulp & Paper Mill; Iron & Steel | 2004 - 2016 | plant |
| eGRID            | Electric Power Industry (grid-connected) | - | 2004, 2005, 2007, 2009, 2010, 2012, 2014, 2016 | plant |
| EIA              | Electric Power Industry | Electric Utility; Industrial; Commercial | 2013 - 2016 | plant |
| EIA              | Electric Power Industry | Electric Utility; IPP NAICS-22 Non-Cogen; IPP NAICS-22 Cogen; Commercial Cogen; Commercial Non-Cogen; Industrial Cogen; Industrial Non-Cogen | 2004 - 2016 | state |
Table 3. Comparison of 2014 electric utility data for Pennsylvania, Ohio, and West Virginia.

| Reporting system | NOx Emissions (short tons) | Net Generation (TWh) | Heat Input (million mmBTU) |
|------------------|---------------------------|----------------------|---------------------------|
| CEMS             | 3.996                     | 71.0                 | 574                       |
| eGRID            | 4.286                     | 75.9                 | 573                       |
| EIA              | 6.701                     | 75.8                 | 594                       |

Emissions from industrial \(E_{\text{industrial},i,j,t}\), commercial \(E_{\text{commercial},i,j,t}\), and residential \(E_{\text{residential},i,j,t}\) end use are estimated as follows:

\[
E_{\text{industrial},i,j,t} = R_{\text{industrial},\text{state},i,j,t} \cdot U_{\text{state},t} \tag{27}
\]

\[
E_{\text{commercial},i,j,t} = R_{\text{commercial},\text{state},i,j,t} \cdot U_{\text{state},t} \tag{28}
\]

\[
E_{\text{residential},i,j,t} = R_{\text{residential},\text{state},i,j,t} \cdot U_{\text{state},t} \tag{29}
\]

where \(R_{\text{industrial},\text{state},i,j,t}\), \(R_{\text{commercial},\text{state},i,j,t}\), and \(R_{\text{residential},\text{state},i,j,t}\) are the reported (or simulated) emissions. We use reported county-level emissions from the EPA NEI for industrial, commercial, and residential natural gas combustion for years 2008, 2011, and 2014. Industrial and commercial emissions includes natural gas that is combusted by industrial and commercial (and institutional) boilers and internal combustion engines, respectively.\(^{43}\) Residential emissions includes natural gas that is combusted for residential household heating, grills, hot water heating, and dryers. For non-reporting years, we estimate emissions for each county by linearly interpolating between years or extrapolating across years; we use an area-weighting approach to develop estimates at the 36 x 36 km grid cell resolution.

**Additional results**

The following are additional results from the air quality emissions, mortality, and damages modeling.
Table 4. Emissions per unit activity for 2004 and 2016 for preproduction and production processes. Mean emissions per unit activity are provided, in addition to the percent change in emissions from 2004 to 2016. Note that emissions reflect systems-level factors (e.g., fleet turnover of rate, percentage of wells with wellhead compressors). Comparison to unit-level emissions for 2009 reported in Roy et al. (2014).

| Segment / process | NO<sub>x</sub> | VOC | PM<sub>2.5</sub> |
|-------------------|----------------|-----|-----------------|
|                   | 2004 | 2016 | %△ | Roy et al. (2014) | 2004 | 2016 | %△ | Roy et al. (2014) | 2004 | 2016 | %△ | Roy et al. (2014) |
| Preproduction (tons/spud wells) |      |      |     |                 |      |      |     |                 |      |      |     |                 |
| Drilling          | 5.07 | 4.40 | -13%| 4.4             | 0.60 | 0.46 | -23%| 0.5             | 0.32 | 0.24 | -23%| 0.3             |
| Hydraulic fracturing | 1.88 | 1.20 | -36%| 2.2             | 0.23 | 0.08 | -8% | 0.25            | 0.14 | 0.05 | -64%| 0.16            |
| Trucking          | 1.83 | 0.12 | -94%| 6.9             | 0.10 | 0.01 | -1% | 0.4             | 0.02 | 0.00 | -94%| 0.07            |
| Well completion   | -    | -    | -   | -               | 0.79 | 0.08 | -8% | 3.8             | -    | -    | -   | -               |
| Production (tons/producing well) |      |      |     |                 |      |      |     |                 |      |      |     |                 |
| Condensate tanks (tons/bbl) | -    | -    | -   | -               | 0.01 | 0.00 | 0%  | 0.0003          | -    | -    | -   | -               |
| Heaters           | 0.08 | 0.08 | 0%  | 0.0             | 0.00 | 0.00 | 0%  | 0.0             | -    | -    | -   | -               |
| Pneumatics        | -    | -    | -   | -               | 0.00 | 0.00 | 0%  | 0.5             | -    | -    | -   | -               |
| Fugitives         | -    | -    | -   | -               | 0.00 | 0.00 | -2% | 0.2             | -    | -    | -   | -               |
| Wellhead compressors | 0.02 | 0.01 | -53%| 1.1             | 0.01 | 0.00 | 0%  | 0.4             | 0.00 | 0.00 | 0%  | 0.01            |

Table 5. Cumulative air pollution emissions and percent attribution for each segment and process across the supply chain.

| Segment / process | 2004 to 2016 Cumulative Emissions |
|-------------------|-----------------------------------|
|                   | NO<sub>x</sub> | VOC | PM<sub>2.5</sub> |
|                   | Emissions (thousand metric tons) | %   | Emissions (thousand metric tons) | %   | Emissions (thousand metric tons) | %   |
| Preparation       | 92   | 14% | 14             | 10% | 5.3             | 15% |
| Production        | 5    | 1%  | 71             | 51% | 0.0             | 0%  |
| Processing        | 0    | 0%  | 1              | 1%  | 0.0             | 0%  |
| Transmission      | 121  | 18% | 18             | 13% | 3.8             | 11% |
| Distribution      | 4    | 1%  | 1              | 1%  | 0.3             | 1%  |
| End use           | 44    | 67% | 34             | 24% | 25.7            | 73% |
| Total Supply Chain | 670  | 100%| 140            | 100%| 35.1            | 100%|

17
Figure 12. Emissions per unit activity from 2004 to 2016. Mean emissions per unit activity are provided.
Figure 13. Emissions by process/segment over time. Preproduction (A) NO\textsubscript{x}, (B) VOC, and (C) PM\textsubscript{2.5} emissions. Production (D) NO\textsubscript{x}, (E) VOC, and (F) PM\textsubscript{2.5} emissions. End use (G) NO\textsubscript{x}, (H) VOC, and (I) PM\textsubscript{2.5} emissions.
Figure 14. Maps of cumulative emissions for each county for different segments of the supply chain. (A) NOx, (B) PM$_{2.5}$, and (C) VOC emissions for upstream (i.e., preproduction, production) segment. (D) NOx, (E) PM$_{2.5}$, and (F) VOC emissions for midstream (i.e., gathering, processing, transmission, distribution) segment. (G) NOx, (H) PM$_{2.5}$, and (I) VOC emissions for end use (i.e., electric, industrial, commercial, residential) segment.
Figure 15. Maps of annual NOx emissions for each county.

Figure 16. Maps of annual VOC emissions for each county.
Figure 17. Maps of annual PM$_{2.5}$ emissions for each county.

Figure 18. Maps of annual premature mortalities using AP3 and ACS concentration-response relationship.
Figure 19. Maps of cumulative premature mortalities using AP3 and ACS concentration-response relationship for each segment of supply chain.

Table 6. Cumulative premature mortalities and monetized damages from 2004 to 2016. Damages in 2017 USD. Mortalities and damages based. Mean damages are provided, as well as, 95% confidence intervals (in parentheses) reflecting uncertainty in the VSL. Estimates based on mean, all cause relative risk values from fine particulate matter reported in Pope et al. (2002) (ACS Cohort) and Lepeule et al. (2012) (Harvard Six Cities cohort).

| Segment          | Premature Mortality | Damages (billion $) |
|------------------|---------------------|---------------------|
|                  | ACS Cohort Study    | Harvard Six Cities  |
| AP3              |                     |                     |
| Upstream         | 461                 | 942                 |
| Midstream        | 450                 | 921                 |
| End use          | 1360                | 2766                |
| Supply chain     | 2270                | 4629                |
| APSCA            |                     |                     |
| Upstream         | 256                 | 657                 |
| Midstream        | 266                 | 684                 |
| End use          | 686                 | 1763                |
| Supply chain     | 1208                | 3103                |
| InMAP            |                     |                     |
| Upstream         | 240                 | 618                 |
| Midstream        | 248                 | 638                 |
| End use          | 640                 | 1643                |
| Supply chain     | 1129                | 2900                |
Climate change model

Emissions model

According to the U.S. Greenhouse Gas Inventory (GHGI), the U.S. natural gas sector in 2016 was the source of approximately a quarter of all GHG emissions and a third of all energy-related emissions.\textsuperscript{44} Within the natural gas sector, almost 90% of emissions result from downstream combustion, a majority of which are from electric power and industrial end use segments, and the remaining 10% stem from upstream and midstream processes, as shown in Table 7.\textsuperscript{44}

Table 7. U.S. Greenhouse Gas Inventory and life cycle assessment GHG emissions across the natural gas supply chain.\textsuperscript{40,44,45}

| Segment                  | U.S. Greenhouse Gas Inventory | Life Cycle Assessment\textsuperscript{a} |
|--------------------------|-------------------------------|---------------------------------------|
|                         | 2004                          | 2016                                  |
|                          | Emissions (CO\textsubscript{2}eq mmt) | Emissions (CO\textsubscript{2}eq mmt) | Emissions (CO\textsubscript{2}eq / MJ) | % Emissions |
| Preproduction            | 11                            | 1                                     | 1.8                                     | 3%         |
| Production               | 57                            | 57                                    | 9.7                                     | 14%        |
| Gathering                | 31                            | 53                                    | b                                       | b          |
| Processing               | 31                            | 33                                    | 4.3                                     | 6%         |
| Transmission and storage | 37                            | 33                                    | 1.4                                     | 2%         |
| Distribution             | 25                            | 12                                    | 0.8                                     | 1%         |
| End Use                  | 1183                          | 1476                                  | 50                                      | 73%        |
| Total Supply Chain       | 1374                          | 1665                                  | 68                                      | 100%       |

\textsuperscript{a} These are life cycle emissions for domestic use of natural gas.
\textsuperscript{b} Gathering emissions are included in production.

In the following sections, we describe regional and unit-level emissions model formulations and assumptions for each segment. Table 8 is a summary of emissions modeling input parameters, including definitions, values, and data sources.

We account for time-varying parameters (where practicable), such as evolving regulation and changing activity factors. While we account for changing well completion regulation and corresponding practices over time, we do not explicitly differentiate between emission factors pre- and post-implementation of the 2016 NSPS subpart OOOO which limits vented and fugitive methane emissions from production wells, gathering stations, processing facilities, and transmission and storage compressor stations.\textsuperscript{46} For the segments regulated under the NSPS subpart OOOO, we use data and modeling results from large-scale methane measurement studies conducted across the natural gas supply chain between 2013 and 2016\textsuperscript{27,28,47–51}; these studies provide the best available methane loss rates for each segment and thus the most appropriate for representing operating practices over the period of our analysis. With respect to electricity generation, we use annually reported plant-level emissions, which we assume inherently incorporate time-varying regulatory, efficiency, and activity factors.

Unlike for the air quality impact model, we do not develop spatially-resolved source GHG emission estimates given that the species considered are well-mixed GHGs and global forcing per unit of emission are independent of geographic location of emission.\textsuperscript{52} Spatial variability in GHG emission sources is potentially relevant from a regulatory perspective, namely with respect to classifying and identifying the regulated community or emitters for policy design and enforcement (e.g., methane superemitter policy).
Table 8. Emissions modeling parameter values, definitions, and data sources.

| Parameter | Parameter Definition | Parameter Value | Units | Source |
|-----------|----------------------|-----------------|-------|--------|
| $\theta_1$,production | fitted parameter for estimating mean methane emission factor for production | 0.60 (0.44 to 0.81) | - | 53 |
| $\theta_2$,production | fitted parameter for estimating mean methane emission factor for production | 1.4 (1.3 to 1.8) | - | 53 |
| $\sigma$,production | fitted parameter for estimating mean methane emission factor for production | 1.3 (1.1 to 1.6) | - | 53 |
| $\beta$,production | fitted parameter for estimating mean methane emission factor for production | 3.0 (2.6 to 3.2) | - | 53 |
| $\delta$,production | fitted parameter for estimating mean methane emission factor for production | -2.2 (-2.6 to -1.8) | - | 53 |
| $\epsilon$,production | fitted parameter for estimating mean methane emission factor for production | 0.20 (0.050 to 0.42) | - | 53 |
| $C_{\text{distribution,CH}_4}$ | distribution methane content | 93.4% | % | 54 |
| $C_{\text{production,CH}_4,t}$ | production methane content for northeast National Energy Modeling System (NEMS) region | time-varying (ranges from 83 to 84%) | % | 54 |
| $C_{\text{production,CO}_2,t}$ | production carbon dioxide content for unconventional natural gas in northeast NEMS region | 3.5% | % | 54 |
| $C_{\text{transmission,CH}_4}$ | transmission and storage methane content | 93.4% | % | 54 |
| $C_{\text{processing,after,CO}_2}$ | carbon dioxide content after processing | 1% | % | 54 |
| $C_{\text{processing,before,CO}_2}$ | carbon dioxide content prior to processing | 7.4% | % | 54 |
| $EF_{\text{combustion}}$ | carbon dioxide emission factor for natural gas combustion | 54 | metric ton/mmcf | 55,56 |
| $EF_{\text{completion,CH}_4,\text{controlled}}$ | methane emissions factor for completions with emission controls | 2.23 (base), 0.66 (low), 4.65 (high) | metric ton/spud well | 27 |
| $EF_{\text{completion,CH}_4,\text{uncontrolled}}$ | methane emissions factor for completions without emission controls | 0.84 (base), 0.38 (low), 1.63 (high) | metric ton/spud well | 27 |
| $EF_{\text{drilling}}$ | carbon dioxide emission factor for drilling | 390 (base), 280 (low), 500 (high) | metric ton/spud well | 40 |
| $EF_{\text{hydraulic}}$ | carbon dioxide emission factor for hydraulic fracturing pumping | 460 (base), 230 (low), 690 (high) | metric ton/spud well | 40 |
| $EF_{\text{main,typ}}$ | methane emissions factors for each mile of main | cast iron: 1.16 (4.31), unprotected steel: 0.86 (2.32), protected steel: 0.10 (0.37), plastic: 0.03 (0.06) | metric ton/mile | 44,50 |
| $EF_{\text{meters,typ}}$ | methane emissions factor for distribution meters and regulators | | metric ton/meter | 44,50 |
| $EF_{\text{service,typ}}$ | methane emissions factor for each service line | unprotected steel: 0.01 (0.04), protected steel: 0.001 (0.002), plastic: 0.0003 (0.0004), copper: 0.005 | metric ton/service line | 44,50 |
| $EF_{\text{wellpad,preparation}}$ | carbon dioxide emission factor for well pad preparation | 340 (base), 300 (low), 360 (high) | metric ton/well | 40 |
| $F_{\text{completion,controlled},\delta}$ | fraction of well completions with emission controls | time-varying | -/- | 27,28 |
| $HHV_{\text{reference}}$ | reference higher heating value used in AP-42 natural gas combustion carbon dioxide emission factors | 1000 | BTU/ft³ | 44 |
| $HHV_{s,t}$ | higher heating value for delivered natural gas | time-varying | BTU/ft³ | 57 |
Table 8 (continued). Emissions modeling parameter values, definitions, and data sources.

| Parameter           | Parameter Definition                                                                 | Parameter Value                  | Units       | Source |
|---------------------|--------------------------------------------------------------------------------------|----------------------------------|-------------|--------|
| $h_{\text{gathering}}$ | methane loss rate for gathering facilities                                             | 0.40% (base), 0.36% (low), 0.45% (high) | %           | 58     |
| $h_{\text{processing}}$ | methane loss rate for processing segment                                              | 0.18% (base), 0.16% (low), 0.20% (high) | %           | 58     |
| $h_{\text{transmission}}$ | methane loss rate for transmission and storage segment                                 | 0.35% (base), 0.28% (low), 0.45% (high) | %           | 47     |
| $M_{\text{distribution,s,t,type}}$ | number of meters and regulators                                                      | time-varying (see Figure 20)     | meters      | 44, 59 |
| $P_{\text{main,s,t,type}}$ | length of main pipeline of each type                                                  | time-varying (see Figure 20)     | miles       | 59     |
| $P_{\text{production,well,t}}$ | production per well                                                                    | time-varying                      | mcf         | 4-6    |
| $P_{\text{service,s,t,type}}$ | number of service lines of each type                                                  | time-varying (see Figure 20)     | service lines | 59   |
| $U_{\text{t}}$ | fraction total unconventional production out of total production                    | time-varying (ranges from 0 to 99%) |/%           | 4-7    |
| $V_{\text{commercial,s,t}}$ | volume of natural gas delivered to commercial consumers                               | time-varying (see Figure 8)      | mmcf        | 15     |
| $V_{\text{completion}}$ | natural gas combusted during well completions for wells with emission controls     | 1.57 (base), 0.46 (low), 2.87 (high) | mmcf/well   | 27     |
| $V_{\text{distribution,s,t}}$ | volume of natural gas delivered to consumers                                          | time-varying (see Figure 8)      | mmcf        | 60     |
| $V_{\text{industrial,s,t}}$ | volume of natural gas delivered to industrial consumers                               | time-varying (see Figure 8)      | mmcf        | 14     |
| $V_{\text{leasefuel,s,t}}$ | natural gas lease fuel consumption                                                     | time-varying (see Figure 8)      | mmcf        | 10     |
| $V_{\text{plantfuel,s,t}}$ | natural gas plant fuel consumption                                                     | time-varying (see Figure 8)      | mmcf        | 11     |
| $V_{\text{processing,s,t}}$ | natural gas processing volume (i.e., unprocessed volume received by plant)            | time-varying (see Figure 8)      | mmcf        | 9      |
| $V_{\text{production,s,t}}$ | shale gas production volume                                                           | time-varying                      | mmcf        | 4-6    |
| $V_{\text{residential,s,t}}$ | volume of natural gas delivered to residential consumers                               | time-varying (see Figure 8)      | mmcf        | 17     |
| $V_{\text{transmissionfuel,s,t}}$ | transmission and distribution fuel consumption                                         | time-varying (see Figure 8)      | mmcf        | 12     |
| $W_{\text{producing,t}}$ | number of producing wells                                                             | time-varying                      | wells       | 4-6    |
| $W_{\text{spud,t}}$ | number of spud wells                                                                  | time-varying                      | wells       | 4-6    |

**Preproduction**

We estimate well pad preparation emissions ($E_{\text{wellpadpreparation,t}}$) from land clearing and well pad construction for each year $t$ as follows:

$$E_{\text{wellpadpreparation,t}} = \sum_s E_{\text{wellpadpreparation}} \cdot W_{\text{spud,s,t}}$$  \hspace{1cm} (30)

where $W_{\text{spud,s,t}}$ is the number of spud wells for each year $t$ and state $s$. We similarly estimate drilling emissions ($E_{\text{drilling,t}}$) as follows:

$$E_{\text{drilling,t}} = \sum_s E_{\text{drilling}} \cdot W_{\text{spud,s,t}}$$  \hspace{1cm} (31)

We also estimate hydraulic fracturing emissions ($E_{\text{hydraulic,t}}$) from pumping as follows:

$$E_{\text{hydraulic,t}} = \sum_s E_{\text{hydraulic}} \cdot W_{\text{spud,s,t}}$$  \hspace{1cm} (32)

We use a range of emission factors for well pad preparation ($E_{\text{wellpadpreparation}}$), drilling ($E_{\text{drilling}}$), and hydraulic fracturing ($E_{\text{hydraulic}}$) based on modeling results from Jiang et al. (2011). We do not account for changing practices and operating efficiencies related to well pad preparation, drilling, and hydraulic fracturing (e.g., a reduction in drilling time, a reduction in fracturing stages, and an increase in the number of wells per pad), which collectively are likely to have trivial impacts on total GHG emissions.
Emissions from well completions are highly uncertain, but are a relatively minor source, contributing less than 1% of life cycle GHG emissions.\textsuperscript{40–42} We estimate methane losses from well completion ($E_{\text{completion,CH}_4}$) as follows:

$$E_{\text{completion,CH}_4,t} = \sum_s [F_{\text{completion,controlled},t} \cdot EF_{\text{completion,CH}_4,\text{controlled}} + (1 - F_{\text{completion,controlled},t}) \cdot EF_{\text{completion,CH}_4,\text{uncontrolled}}] \cdot W_{\text{spud},s,t}$$

We attempt to reflect voluntary adoption of emission controls and rapidly evolving regulation, including the implementation of the NSPS subpart OOOO and National Emissions Standards for Hazardous Air Pollutants (NESHAP) standards, that require reduced emission completions (RECs) for hydraulically fractured wells by 2015.

We additionally derive methane emission factors for wells with emission controls ($EF_{\text{completion,CH}_4,\text{controlled}}$), and without emission controls ($EF_{\text{completion,CH}_4,\text{uncontrolled}}$) based on a measurement study conducted by Allen et al. (2013).\textsuperscript{27} Given insufficient sample size, we do not further segregate types of emission controls. To derive a range of emission factor estimates, we employ a bootstrapping method used in other emission studies\textsuperscript{27,61–63}; we resample with replacement $n$ times from the dataset (where $n=25$ is the sample size), estimate the mean of the bootstrapped sample, iterate 100,000 times, and then find the mean and 95% confidence interval across the bootstrapped means.

We assume a changing fraction of well completions with emission controls ($F_{\text{completion,controlled},t}$) and without emission controls ($F_{\text{completion,uncontrolled},t}$). The penetration of emission controls over time is uncertain, although recent emission measurement studies suggest that most wells in Appalachia (for which measurements were taken) have emission controls.\textsuperscript{27,28}

We also estimate carbon dioxide flaring emissions from well completion ($E_{\text{completion,CO}_2}$) as follows:

$$E_{\text{completion,CO}_2,t} = \sum_s F_{\text{completion,controlled},t} \cdot V_{\text{completion}} \cdot EF_{\text{combustion}} \cdot W_{\text{spud},s,t}$$

where $EF_{\text{combustion}}$ is the carbon dioxide emissions factor for natural gas combustion. The volume of natural gas combusted during controlled well completions ($V_{\text{completion}}$) is derived using the previously described bootstrapping method, using data from Allen et al. (2013).\textsuperscript{27} Although we do not explicitly assume the flaring rate as in several previous studies, the combusted volume incorporates some observations in which flaring is employed.\textsuperscript{40–42,64}

Production

We develop a bottom-up estimate of methane losses from producing wells ($E_{\text{production},s,t}$) based on a range of emissions factors ($EF_{\text{production,CH}_4}$) conditional on site-level production as follows:

$$E_{\text{production,CH}_4,t} = \sum_{\text{well}} EF_{\text{production,CH}_4,\text{well},t} \cdot W_{\text{producing},t}$$

where the emission factor is:

$$EF_{\text{production,CH}_4,\text{well},t} = e^{\mu_{\text{production,well},t} + \frac{1}{2}\sigma_{\text{production},t}^2}$$

where $\sigma_{\text{production}}$ is the standard deviation (fitted value) and $\mu_{\text{production,well},t}$ is the mean given by:

$$\mu_{\text{production,well},t} = a_{\text{production}} + b_{\text{production}} \cdot P_{\text{production,well},t}^{\theta_1_{\text{production}}} + c_{\text{production}} \cdot P_{\text{production,well},t}^{\theta_2_{\text{production}}}$$

where $a_{\text{production}}$, $b_{\text{production}}$, $\theta_1_{\text{production}}$, and $\theta_2_{\text{production}}$ are fitted parameters, and $P_{\text{production,well},t}$ is the annual production per well. The empirical relationship defining the emission
factor conditional on site-level production described by equations 36 and 37 is that derived in Alvarez et al. (2018), a study which comprehensively evaluated methane emissions for production based on several recent measurement studies. The study provides a range of fitted values for the parameters specific for the basin, defining the mean and 95% confidence interval of emission factors. While the empirical relationship is based on site-level rather than well-level data, we assume that the relationship applies at the well level; based on a GIS cluster analysis defining the number of wells per site in Alvarez et al. (2018), 96% of sites within the basin have one well.

We also estimate carbon dioxide emissions from combustion of lease fuel \( (E_{\text{production}, \text{CO}_2, t}) \) as follows:

\[
E_{\text{production}, \text{CO}_2, t} = \sum_s V_{\text{leasefuel}, s, t} \cdot U_{s, t} \cdot EF_{\text{combustion}}
\]  

(38)

where \( V_{\text{leasefuel}, s, t} \) is the lease fuel consumed (as reported by the EIA) and \( U_{s, t} \) is the unconventional prorate factor.

Gathering

We develop a top-down estimate of methane losses from gathering facilities \( (E_{\text{gathering}, s, t}) \) based on gathering volume and methane loss rate as follows:

\[
E_{\text{gathering}, t} = \sum_s (V_{\text{production}, s, t} - V_{\text{leasefuel}, s, t} \cdot U_{s, t}) \cdot C_{\text{production}, t} \cdot L_{\text{gathering}}
\]  

(39)

where \( V_{\text{production}, s, t} \) is the shale production volume. We develop scenarios by varying the main source of uncertainty, the methane loss rate \( (L_{\text{gathering}}) \). We use loss rate estimates from a study by Marchese et al. (2015), which is based on recent measurements of gathering facility emissions across the U.S., and specifically use loss rate estimates based on measurements in Pennsylvania, given that there is very high variability across regions; we further adjust central and high estimates to account for heavy-tailed distributions, as described in Alvarez et al. (2018). We do not model gathering pipeline leaks, given outdated and otherwise insufficient activity and emissions data. A recent study conducted by Zimmerle et al. (2017) measures gathering pipeline leaks, and while the study suggests that the GHG may underestimate gathering pipeline leaks, study data were reported as insufficiently representative to develop emission factors.

Processing

We estimate fugitive and vented methane losses \( (E_{\text{processing}, \text{CH}_4, t}) \) as follows:

\[
E_{\text{processing}, \text{CH}_4, t} = \sum_s V_{\text{processing}, s, t} \cdot U_{s, t} \cdot C_{\text{production}, t} \cdot L_{\text{processing}}
\]  

(40)

where \( V_{\text{processing}, s, t} \) is the natural gas volume received by processing plants (as reported for each state by the EIA) and \( C_{\text{production}, t} \) is the methane content of natural gas prior to processing. We develop scenarios by varying the main source of uncertainty, the methane loss rate \( (L_{\text{processing}}) \). We use loss rate estimates from a study by Marchese et al. (2015), which is based on recent measurements of gathering facility emissions across the U.S.; we further adjust central and high estimates to account for heavy-tailed distributions, as described in Alvarez et al. (2018).

We estimate carbon dioxide vented \( (E_{\text{processing}, \text{CO}_2 \text{ venting}, t}) \) for each year \( t \) as follows:

\[
E_{\text{processing}, \text{CO}_2 \text{ venting}, t} = \sum_s V_{\text{processing}, s, t} \cdot U_{s, t} \cdot (C_{\text{B processing}} - C_{\text{A processing}})
\]  

(41)

where \( C_{\text{B processing}} \) is the carbon dioxide content prior to venting and \( C_{\text{A processing}} \) is the carbon dioxide content after venting to achieve a transmission grade composition of natural gas.
We also estimate carbon dioxide emissions from combustion of plant fuel \( E_{\text{processing.CO2 combustion},s,t} \) as follows:

\[
E_{\text{processing.CO2 combustion},s,t} = \sum_s V_{\text{plantfuel},s,t} \cdot U_{s,t} \cdot CE
\]

where \( V_{\text{plantfuel},s,t} \) is the plant fuel consumed.

**Transmission and storage**

We develop a top-down estimate of methane losses from transmission and storage infrastructure \( E_{\text{transmission.CH4},s,t} \) as follows:

\[
E_{\text{transmission.CH4},s,t} = \sum_s (V_{\text{distribution},s,t} + V_{\text{transmissionfuel},s,t}) \cdot U_{s,t} \cdot C_{\text{transmission}} \cdot L_{\text{transmission}}
\]

where \( C_{\text{transmission}} \) is the methane content of transmission natural gas. Consistent with assumptions in Tong et al. (2015), we assume that the annual transmission volume is the sum of pipeline and distribution fuel use \( (V_{\text{distribution},s,t}) \) and volume delivered to end use customers \( (V_{\text{transmissionfuel},s,t}) \) (as reported by the EIA for each state). We model a range of methane loss rate \( (L_{\text{transmission}}) \) scenarios; we use methane loss rates for the entire transmission and storage segment derived in a study by Zimmerle et al. (2015) that combines recent measurements across the U.S. with pipeline losses estimated in the GHGI.

We also estimate carbon dioxide emissions from combustion of transmission and distribution fuel \( E_{\text{transmission.CO2},s,t} \) as follows:

\[
E_{\text{transmission.CO2},s,t} = \sum_s V_{\text{transmissionfuel},s,t} \cdot U_{s,t} \cdot EF_{\text{combustion}}
\]

Given that the EIA reports transmission and distribution fuel consumption \( (V_{\text{transmissionfuel},s,t}) \) in aggregate, we combine combustion emission estimates across segments.

**Distribution**

We estimate methane losses from distribution infrastructure \( E_{\text{distribution.CH4},s,t} \) as follows:

\[
E_{\text{distribution.CH4},s,t} = \sum_s \sum_{\text{type}} (P_{\text{main},s,t,\text{type}} \cdot EF_{\text{main},\text{type}} + P_{\text{service},s,t,\text{type}} \cdot EF_{\text{service},\text{type}} + M_{\text{distribution},s,t,\text{type}} \cdot EF_{\text{distribution},\text{type}}) \cdot U_{s,t}
\]

where \( P_{\text{main},s,t,\text{type}} \) is the miles of mains of each material type and \( P_{\text{service},s,t,\text{type}} \) is the number of service lines of each material type, based on annual data reported by distribution operators to the Pipeline and Hazardous Materials Safety Administration (PHMSA). \( M_{\text{distribution},s,t,\text{type}} \) is the number of meters and regulators of each type. We employ emission factors \( (EF_{\text{main},\text{type}}, EF_{\text{service},\text{type}}, \text{and } EF_{\text{distribution},\text{type}}) \) reported in the GHGI and in a measurement study by Lamb et al. (2015) of 13 urban distribution systems; we model base and high scenarios\(^\text{44,50}\). Rather than use the distribution system loss rate (0.10% to 0.22%) derived in Lamb et al. (2015), which accounts for the distribution of pipelines and meters of each type across the U.S., we use state-specific PHMSA activity data and emission factors by material type over time to reflect recent pipeline replacement efforts in urban areas, such as Pittsburgh, Pennsylvania, and Cincinnati, Ohio. The implied methane loss rate under base case assumptions is 0.21% in 2004 and decreases to 0.15% in 2016.
End Use

We consider emissions from electric power generation, industrial, commercial, and residential end uses.

We use the state-level EIA emissions dataset because it provides estimates for all years from 2004 to 2016, includes emissions from all facilities within the electric power sector, and allows for sub-sector segregation to prevent double-counting. We estimate annual emissions from electric power generation from utilities \( (E_{\text{electric utilities},s,t}) \) as follows:

\[
E_{\text{electric utilities},s,t} = \sum_s RE_{\text{electric utilities},s,t} \cdot U_{s,t}
\]  

where \( RE_{\text{electric utilities},s,t} \) is the reported emissions. Electric power generation from large industrial and commercial facilities are included in emission estimates for those segments.

We estimate annual emissions from industrial facilities \( (E_{\text{industrial},s,t}) \) as follows:

\[
E_{\text{industrial},s,t} = \sum_s CE \cdot V_{\text{industrial},s,t} \cdot \frac{HHV_{s,t}}{HHV_{\text{reference}}} \cdot U_{s,t}
\]  

where \( V_{\text{industrial},s,t} \) is the volume of industrial consumption of natural gas, \( HHV_{s,t} \) is the higher heating value for each state \( s \) and year \( t \), and \( HHV_{\text{reference}} \) is the reference higher heating value.

While the EIA reports emissions for a subset of large industrial facilities comprising less than 1% of industrial fuel consumption, we use the preceding estimation approach for all facilities. We use similar estimation methods for commercial and residential consumption.

Climate impact model

We use the metric, global temperature change, to assess the temporal trace and cumulative impact of natural gas activity on climate. We additionally monetize the impacts based on the social cost of carbon and methane.

We use absolute rather than relative metrics that normalize impacts across species to a reference gas, such as global warming potential (GWP) and global temperature potential (GTP). GWP, the time-integrated radiative forcing due to a pulse emission relative to that of CO\(_2\), has become a default metric because of its simplicity, despite well-documented criticisms of its formulation. GWPs may be inappropriate for evaluating long-term effects, given that they do not take into account that if radiative forcing is applied for a short period, the climate system has time to relax back to equilibrium. A related metric that incorporates additional physical processes and has a less ambiguous interpretation, GTP, is the ratio of a change in global mean surface temperature at a
moment in time in response to an emission pulse relative to that of CO$_2$. While these metrics lend themselves to a benefit-cost or cost-effectiveness framing of climate policy decisions, they do not facilitate understanding the temporal trace of impacts.

We focus on global temperature change, whereby the relationship between the equilibrium global mean surface temperature response ($\Delta T$) and sustained radiative forcing ($RF$) has the general form:

$$\Delta T = \lambda \cdot RF$$ \hspace{1cm} (48)

where $\lambda$ is the climate sensitivity parameter. Radiative forcing, typically expressed in units of watts per square meter, is a commonly used metric describing the net change in the energy balance of the climate system resulting from an imposed perturbation, such as emissions of GHGs from the natural gas sector. Often radiative forcing and metrics derivative of radiative forcing are proportional and describe the relation to the global mean temperature response, with fewer metrics describing the relation to other climate change phenomena. Furthermore, radiative forcing-based metrics do not incorporate climate effects unrelated to radiative forcing, such as the effects of land cover change on evapotranspiration. Thus, while there is utility in quantifying the global mean temperature change resultant of an emission, such an exercise is imperfect given its limited perspective on the factors driving broader climate change.

An explicit form of the relationship is given by the following convolution of an emission scenario and the average global temperature potential (AGTP):

$$\Delta T(t) = \sum_i \int_0^t E_i(s) \cdot AGTP_i(t - s) \, ds$$ \hspace{1cm} (49)

where $s$ is the time of an emissions pulse and $t$ is the time of an emissions response. $E_i(s)$ for species $i$ is an emissions scenario (which we formulated and described in Section 0). AGTP is the temperature change at time $t$ due to 1-kg emission at $t = 0$, typically expressed in units of K kg$^{-1}$. The general form of AGTP is given by:

$$AGTP_i(t) = \int_0^t RF_i(s) \cdot R_T(t - s) \, ds$$ \hspace{1cm} (50)

where $RF_i$ is the radiative forcing due to a pulse emission and $R_T$ is the temperature response to a unit of forcing, both of which are parameterized based on more complex models that explicitly include physical processes. For well-mixed greenhouse gases, the general form of radiative forcing has the form:

$$RF_i = A_i \cdot \exp \left( - \frac{t}{\tau_i} \right)$$ \hspace{1cm} (51)

where $\tau$ is the perturbation lifetime for the removal of the gas from the atmosphere and $A_i$ is the radiative forcing per unit mass increase in atmospheric concentration (i.e., radiative efficiency). It is assumed that $A_i$ and $\tau_i$ are independent of the concentration of greenhouse gases; in reality, there are dependencies and nonlinearities in $A_i$ and $\tau_i$ which can lead to systematic biases in the absolute value of these metrics. The climate response function $R_T$ is given by:

$$R_T(t) = \sum_{j=1}^M c_j / d_j \cdot \exp \left( - \frac{t}{d_j} \right)$$ \hspace{1cm} (52)

where $c_j$ are climate sensitivity parameters and $d_j$ are response time parameters for all terms $j = 1, \ldots, M$. The absolute global temperature potentials for methane and carbon dioxide are given by.
\[
AGTP_{CH4}(t) = (1 + f_1 + f_2)A_{CH4} \sum_{j=1}^{2} \frac{\tau_{CH4} c_j}{\tau_{CH4} - d_j} \left[ \exp\left(-\frac{t}{\tau_{CH4}}\right) - \exp\left(-\frac{t}{d_j}\right) \right]
\] (53)

\[
AGTP_{CO2}(t) = A_{CO2} \sum_{j=1}^{2} \left( a_0 c_j \right) \left[ 1 - \exp\left(-\frac{t}{d_j}\right) \right] + \sum_{k=1}^{3} \frac{a_k \tau_{CO2,k} c_j}{\tau_{CO2,k} - d_j} \left[ \exp\left(-\frac{t}{\tau_{CO2,k}}\right) - \exp\left(-\frac{t}{d_j}\right) \right]
\] (54)

where there are multiple exponential terms \( j \). \( a_k \) for terms \( k = 0, ..., 3 \) are coefficients describing the fraction of CO\(_2\) remaining in the atmosphere after a pulse. The additional terms \( f_1 \) and \( f_2 \) are effects due to ozone and stratospheric water, respectively. The radiative efficiency of carbon dioxide and methane is given by\(^{52}\):

\[
A_{CO2} = \alpha \left[ \frac{\log(C_{0,t} + \Delta C)}{\Delta C} \right]
\] (55)

where \( \alpha \) is the radiative transfer coefficient. \( C_{0,t} \) is the reference CO\(_2\) concentrations and \( \Delta C \) is the change from the reference concentration (which we evaluate at \( \Delta C = 1 \) ppm\(_v\)). For emission pulses in years 2004 to 2016, we use observed CO\(_2\) global atmospheric concentrations reported by the National Oceanic and Atmospheric Administration (NOAA), and for pulses after 2016, we use CO\(_2\) concentrations for four IPCC RCP scenarios. The AGTP functions for 1-kg pulses from years 2004 to 2016 are depicted in Figure 21. We use a Monte Carlo simulation approach to reflect uncertainty in the AGTP values; we assign probability distributions to key input parameters and iterate 10,000 times. The mean and 95% confidence interval of the AGTP are provide in Figure 21. Parameters, including their definitions, values or distributions, and sources, are provided in Table 9.

Table 9. Climate impact input parameters, definitions, values, units, and sources.

| Parameter | Parameter Definition | Parameter Value / Distribution\(^a\) | Units | Source |
|-----------|----------------------|--------------------------------------|-------|--------|
| \( \alpha \) | radiative transfer coefficient | 5.35 | W/m\(^2\) | \(^{52}\) |
| \( A_{CH4} \) | radiative efficiency of CH\(_4\) | 3.62 x 10\(^{-4}\) | W/m\(^2\)/ppb\(_v\) | \(^{52}\) |
| \( a_k \) | fraction of CO\(_2\) remaining in the atmosphere after a pulse | \( k = 0: 0.2173 \)
\( k = 1: 0.2240 \)
\( k = 2: 0.2824 \)
\( k = 3: 0.2763 \) | unitless | \(^{74}\) |
| \( C_{0,t} \) | reference mean atmospheric CO\(_2\) concentration* | time-varying | ppm\(_v\) | \(^{76}\) |
| \( c_j \) | climate sensitivity parameters | \( j = 1: \) Uniform(0.631±0.02) \( j = 2: \) Uniform(0.429±0.18) | K / W / m\(^2\) | \(^{67,77}\) |
| \( d_j \) | response time parameters | \( j = 1: \) Uniform(8.4±30%) \( j = 2: \) Uniform(409.5±30%) | year | \(^{67,77}\) |
| \( f_1 \) | ozone effect on CH\(_4\) radiative forcing | Normal(0.5, 0.05) | | \(^{52,75,77}\) |
| \( f_2 \) | stratospheric water effect on CH\(_4\) radiative forcing | Normal(0.15, 0.05) | | \(^{75,77}\) |
| \( \tau_{CH4} \) | CH\(_4\) perturbation lifetime | Normal(12.4, 1) | years | \(^{67,77}\) |
| \( \tau_{CO2,k} \) | CO\(_2\) perturbation lifetime for each term | \( \tau_{k=1} = 394.4 \)
\( \tau_{k=2} = 36.54 \)
\( \tau_{k=3} = 4.304 \) | years | \(^{74}\) |

\(^a\) We represent normal distributions as Normal(mean, standard deviation). We represent uniform distributions as Uniform(base value ± error term), where the minimum is the base value minus error term and the maximum is the base value plus the error term.

\(^b\) We convert from values given in ppm\(_v\) to kg, assuming the mean molecular weight of air is 28.97 kg/kmol, the molecular weight of methane is 16.04 g/mol, and the total mass of the atmosphere is 5.1352 x 10\(^{18}\) kg.

\(^c\) We convert from values given in ppm\(_v\) to kg, assuming the mean molecular weight of air is 28.97 kg/kmol, the molecular weight of methane is 44.01 g/mol, and the total mass of the atmosphere is 5.1352 x 10\(^{18}\) kg.
Figure 21. Absolute global temperature potential from 2004 to 2100 for 1-kg pulses in years 2004 to 2016. (A) CO$_2$ and (B) CH$_4$. Base scenario (based on simulated mean) indicated in green. Low scenario (based on simulated lower 95% confidence interval) indicated in blue. High scenario (based on simulated upper 95% confidence interval) indicated in red.

The physical characteristics of CO$_2$ and CH$_4$, including the relative atmospheric lifetimes, have implications for policy design and evaluation related to the natural gas system.$^{78,79}$ CH$_4$ emissions, which can be treated as a flow, can be addressed through policies targeting reductions using the existing suite of cost-effective abatement technologies (e.g., replacing leaking components at compressor stations). CH$_4$ policies can reduce near-term warming rates resulting in climate benefits realized over relatively short time horizons, and may also be key in the avoidance or delay of reaching “tipping points” in the climate system, irreversible thresholds with drastic consequences.$^{80}$ In contrast, CO$_2$ emissions, which can be treated as a stock, can only be achieved through more systemic policies focused on deep decarbonization or a fundamental transition of the energy system away from fossil fuels, including natural gas, with the benefits of such policies largely derived by future generations.

Additional results

Estimated emissions and the percent attribution of emissions across the supply chain are reasonably consistent with other studies.
Table 10 provides cumulative emissions by supply chain segment and process and the percent attribution for each segment. Balcombe et al. (2017) find that CH₄ accounts for 53% of GHG emissions (normalized in terms of CO₂eq) across the supply chain (excluding end use); we similarly find that CH₄ accounts for 66% (57-73%) of supply chain emissions (excluding end use) or 23% (16-30%) of supply chain emissions (including end use). Table 11 includes a comparison between this study, and the Alvarez et al. (2018) study and 2015 US GHGI inventory. The percent attribution of methane emissions across segments between this study and the Alvarez et al. (2018) study are similar, with a majority of emissions associated with production. However, the methane loss rates estimated in this study, although similar to the US GHGI and the source-based estimates in the Alvarez study, are lower than the site-based estimates in the Alvarez study. This may be attributed to differences in source inclusion, estimation methods, regional emission factors, and study scope (i.e., Appalachian basin versus U.S., shale gas versus O&G sector). Although we estimate a declining loss rate over time, this trend is largely a byproduct of estimation methods and increasing well productivity (Figure 22).

Figure 22. Methane loss rates over time. Blue, orange, and gray bars represent base, low, and high loss rate estimates, respectively, based on emission scenario estimates. Loss rates assume state-reported production and time-varying CH₄ content (~83-84%). Yellow line represents production over time.
Table 10. Cumulative emissions and percent attribution for each segment and process across the supply chain. Base scenario emission estimates with low and high scenario estimates provided in the parentheses.

| Segment / process          | 2004 to 2016 Cumulative Emissions | CO₂ Emissions (mmt) | %  | CH₄ Emissions (mmt) | %  |
|----------------------------|-----------------------------------|---------------------|----|--------------------|----|
| Preproduction              | 15.5 (10.2 – 20.6)                | 2%                  |    | 0.02 (0.01 - 0.05) | 0% |
| Well pad preparation       | 4.17 (3.68 - 4.42)                | 1%                  |    | -                  |    |
| Drilling                   | 4.79 (3.44 - 6.14)                | 1%                  |    | -                  |    |
| Hydraulic fracturing       | 5.65 (2.82 - 8.46)                | 1%                  |    | -                  |    |
| Well completion            | 0.88 (0.26 - 1.61)                | 0%                  |    | 0.02 (0.01 - 0.05) | 0% |
| Production                 | 50.0                              | 7%                  | 3.69 (2.10 – 5.73) | 63% |
| Lease fuel consumption     | 50.0                              | 7%                  |    | -                  |    |
| Fugitive / vented losses   | -                                 | -                   | 3.69 (2.10 – 5.73) | 63% |
| Gathering                  | -                                 | -                   | 0.91 (0.65 - 1.17) | 15% |
| Fugitive / vented losses   | -                                 | -                   | 0.91 (0.65 - 1.17) | 15% |
| Processing                 | 12.2                              | 2%                  | 0.25 (0.22 - 0.27) | 4%  |
| Plant fuel consumption     | 3.45                              | 1%                  |    | -                  |    |
| Fugitive / vented losses   | 8.70                              | 1%                  | 0.25 (0.22 – 0.27) | 4%  |
| Transmission and storage   | 26.0                              | 4%                  | 0.69 (0.55 - 0.89) | 12% |
| Fuel consumption           | 26.0                              | 4%                  |    | -                  |    |
| Facility / pipeline fugitive losses | -                       | -                   | 0.69 (0.55 - 0.89) | 12% |
| Distribution               | -                                 | -                   | 0.30 (0.30 - 0.64) | 5%  |
| Fugitive losses            | -                                 | -                   | 0.30 (0.30 - 0.64) | 5%  |
| End use                    | 571                               | 85%                 |    | -                  |    |
| Electricity generation     | 175                               | 26%                 |    | -                  |    |
| Industrial use             | 146                               | 22%                 |    | -                  |    |
| Commercial use             | 96                                | 14%                 |    | -                  |    |
| Residential use            | 152                               | 23%                 |    | -                  |    |
| Total Supply Chain         | 675 (669 - 680)                   | 100%                | 5.87 (3.83 – 8.75) | 100%|
Table 11. Comparison of 2015 methane emissions, percent emissions, and loss rates across this study, the Alvarez et al. (2008) study\textsuperscript{53}, and the U.S. GHGI.

| Segment              | This work | Alvarez et al. (2018) (site-based)\textsuperscript{a} | Alvarez et al. (2018) (source-based)\textsuperscript{b} | U.S. GHGI  |
|----------------------|-----------|--------------------------------------------------------|--------------------------------------------------------|------------|
|                      |           | Emissions [CH\textsubscript{4} mmt] % Emissions | Emissions [CH\textsubscript{4} mmt] % Emissions | Emissions [CH\textsubscript{4} mmt] % Emissions | Emissions [CH\textsubscript{4} mmt] % Emissions |
| Appalachian basin    |           |                                                      |                                                      |                                                      |                                                      |
| Preproduction        | 0 (0-0.01) 0% | 0.09 (0.08-0.12)\textsuperscript{c} 1% | 0.09 (0.08-0.12)\textsuperscript{c} 1% | 0.10\textsuperscript{c} 1% |
| Production           | 0.76 (0.43-1.18) 61% | 7.2 (5.6-9.1)\textsuperscript{d} 56% | 2.8 (2.7-2.9)\textsuperscript{d} 33% | 3.10\textsuperscript{g} 40% |
| Gathering            | 0.22 (0.16-0.28) 17% | 2.6 (2.4-3.2) 20% | 2.6 (2.4-3.2) 31% | 2.30 30% |
| Processing           | 0.07 (0.07-0.08) 6% | 0.72 (0.65-0.92) 6% | 0.72 (0.65-0.92) 9% | 0.45 6% |
| Transmission and storage | 0.14 (0.11-0.18) 11% | 1.8 (1.6-2.1) 14% | 1.8 (1.6-2.1) 21% | 1.30 17% |
| Distribution         | 0.06 (0.06-0.12) 5% | 0.44 (0.22-0.95) 3% | 0.44 (0.22-0.95) 5% | 0.44 6% |
| Total supply chain   | 1.25 (0.82-1.84) 100% | 12.85 (10.6-16.4) 100% | 8.45 (7.65-10.2) 100% | 7.69 100% |
| Supply chain loss rate (%) | 1.16 (0.76 – 1.71)\textsuperscript{e} | 2.3 (2.0 - 2.7)\textsuperscript{f} | 1.48 (1.34 – 1.78)\textsuperscript{g} | 1.34\textsuperscript{g} |

Most values are bottom-up, site-based estimates derived in Alvarez et al. (2018). They include emissions across the oil and natural gas supply chain and are not exclusive of shale gas.

\textsuperscript{a} Values are bottom-up, site-based estimates derived in Alvarez et al. (2018). They include emissions across the oil and natural gas supply chain and are not exclusive of shale gas.

\textsuperscript{b} Includes emissions from both completions and workovers.

\textsuperscript{c} Includes emissions from routine operations.

\textsuperscript{d} Loss rate is estimated as a percentage of methane produced. Assumes state-reported shale production (refer to section 2.2), and time-varying methane content for the Northeast (~83-84%).

\textsuperscript{e} Reported loss rate, as a function of total methane produced (33 tcf, with average CH\textsubscript{4} content of 90 vol%), derived in Alvarez et al. (2018). The range represents the reported 95% confidence interval. That study also reported a loss rate of 2.9%, as a percentage of total methane delivered (25 tcf/y NG delivered, assuming an average CH\textsubscript{4} content in NG of 95% by volume).

\textsuperscript{f} Estimated loss rate, as a function of total methane produced (33 tcf, with average CH\textsubscript{4} content of 90 vol%), similar to the approach used in Alvarez et al. (2018).

Table 12. Cumulative climate change temperature impacts.

| Time-Integrated Cumulative Temperature Impact from 2004 (Kelvin-years) |  |
|-------------------------------------------------|-----------------|
| to 2016                                         | 0.001 (0-0.001) |
| to 2050                                         | 0.023 (0.014-0.037) |
| to 2100                                         | 0.045 (0.029-0.067) |
| to 2200                                         | 0.082 (0.055-0.116) |

| Instantaneous Temperature Impact (Kelvin) |     |
|------------------------------------------|-----|
| 2016                                     | 0.0003 (0.0002-0.0006) |
| 2050                                     | 0.0005 (0.0004-0.0008) |
| 2100                                     | 0.0004 (0.0003-0.0005) |
| 2200                                     | 0.0004 (0.0003-0.0005) |
Figure 23. Annual temperature impact from sources within Appalachia indicating contributions from each segment of the supply chain. Dotted black lines depict temperature impact under low and high scenarios.

Figure 24. Annual temperature impact from sources within Appalachia indicating contributions from each year of natural gas activity. Dotted black lines depict temperature impact under low and high scenarios.
Employment model

Data

Employment and Earnings

We use the Bureau of Economic Analysis (BEA) Local Area Personal Income and Employment (LAPI) datasets, which provide county-level estimates from 2005 to 2015. The datasets are mainly derived from administrative records data of various federal and state government social insurance programs and tax codes. These data originate from the recipients of the income or from the payer of the income.

We use total employment estimates, which consist of wage and salary employment and proprietor’s income. Wage and salary employment, as defined by the BEA, measures the average annual number of full- and part-time jobs in each area by place of work, including all jobs for which wages and salaries are paid.

We also estimate the share of total earnings from different sectors (i.e., farm, construction, manufacturing, retail, and mining), based on total and sector-level earnings. Earnings, as defined by the BEA, consist of compensation of employees and proprietors’ income.

Population and Population Density

We use the BEA estimates of county-level population, and combining those data with county-level land area values from the Census Bureau, we estimate population density.

Employment rate

We estimate the employment rate, which is the number of employed divided by the total labor force. The number employed and total labor force, as reported by the Bureau of Labor Statistics (BLS) Local Area Unemployment Statistics, are based on the Community Population Survey and unemployment insurance disaggregated from state-level statistics to the county-level; the processing of this data may not accurately reflect the county-level employed and labor force, thus limiting its utility for regression. The total labor force are those persons greater 16 years old, and the number employed are any persons that have worked (not necessarily full-time) or are on leave, vacation, illness, etc. during the survey.

Employment-to-population ratio

The employment-to-population ratio is the total employment divided by the population (as reported in BEA LAPI). In some instances, the employment-to-population ratio is greater than, demonstrating the limitations of the underlying data.

Shale counties

We identify shale counties based on United States Geological Service (USGS) designations for Marcellus and Utica shale subunits. If at least 10% of a county’s area is within the footprint of the shale subunits, we classify it as a shale county. Note that several producing shale formations other than the Marcellus and Utica exist within the Appalachian basin; however, they are largely within the footprint of the Marcellus and Utica formations, thus, we do not identify them separately. The shale classifier is used to subset the panel dataset for model fitting. Several other empirical studies, including Parades (2015), partition or exclude counties which are not within the Marcellus (or respective shale formation).
Nonmetropolitan counties

We identify nonmetropolitan counties based on a cross-county comparison of 2010 population estimates, whereby we classify counties as metropolitan if they are within the top 10 percentile based on population. Of the 272 counties within Pennsylvania, Ohio, West Virginia, and New York, 28 counties are classified as metropolitan, two of which produced (including Allegheny County, Pennsylvania, which had 49 producing wells in 2015, and Stark County, Ohio, which had only 2 producing wells in 2015). The metropolitan classifier is used to subset the panel dataset for model fitting. Including only nonmetropolitan counties creates a more homogenous sample, precluding counties with large cities from excessively influencing estimates from a linear model (Weber 2014). Moreover, nonmetropolitan counties with thin labor markets are presumably the population of interest.

Modeling specification

County-level descriptive statistics of the panel data are provided in Table 13. We begin with a dataset consisting of all 272 counties within Pennsylvania, Ohio, West Virginia, and New York, of which 93 counties produced unconventional natural gas and 195 counties are within (or partially within) the Marcellus and/or Utica shale plays. The full dataset includes 2972 observations over the period 2005 to 2015, accounting for observations that were removed due to missing data. Sector-level earnings were obscured for some counties in some years, so we exclude these observations. In addition, Pennsylvania production (and resultingly well count) data is unavailable for 2010 due to reporting inconsistencies, further reducing the number of observations. Unless otherwise specified, we use the full dataset to fit the following models.

As part of the model selection process, we begin by fitting and comparing five types of models that can be used with panel datasets: pooled ordinary least squares (OLS), OLS fixed effects, general least squares (GLS) fixed effects, first-difference, and random effects. Using panel data allows us to control for several types of unobserved heterogeneity that could potentially confound the estimated effect of natural gas activity. The first specification measuring the contemporaneous effect of natural gas wells using pooled OLS is as follows:

\[
Y_{ct} = \beta \text{ProducingWells}_{ct} + X_{ct} \theta + \alpha + \varepsilon_{ct}
\]

Where \(Y_{ct}\) is a measure employment for each county \(c\) and year \(t\), and \(\alpha\) is a constant coefficient, assuming no fixed heterogeneity across counties or years. \(\text{ProducingWells}_{ct}\) is the number of producing unconventional natural gas wells, and \(\beta\) can be interpreted as the average change in employment attributable to another producing well in the county. \(X_{ct}\) are labor market and demographic control variables, which are as follows:

\[
X_{ct} = \{\text{population}_{ct}, \text{populationdensity}_{ct}, \text{retailshareearnings}_{ct}, \text{constructionshareearnings}_{ct}, \text{manufacturingshareearnings}_{ct}, \text{retailshareearnings}_{ct}\}
\]

These control variables are similar to those used in Paredes (2015) and Weber (2014). \(\theta\) are the coefficients associated with the control variables. \(\varepsilon_{ct}\) is the random error term given by:

\[
\varepsilon_{ct} = \mu_{c} + \lambda_{t} + \varepsilon_{ct}
\]

where \(\mu_{c}\) is the county error component, \(\lambda_{t}\) is the time error component, and \(\varepsilon_{ct}\) is the idiosyncratic error. The composition of the error term depends on whether the model incorporates county, time, or both types of effects. To test for the significance of county, year, and two-way effects, we use the
Lagrange multiplier test for pooled OLS, finding that county effects are significant, but time effects are not. We note that the average change in employment from a producing well $\beta$ is not significant.

We then specify fixed effects models, including alternatively county, year, and two-way effects, as follows:

$$Y_{ct} = \beta ProducingWells_{ct} + X_{ct}\theta + \alpha_c + \gamma_t + \varepsilon_{ct}$$  \hspace{1cm} (59)

where $\alpha_c$ are county fixed effects and $\gamma_t$ are year fixed effects. The fixed effects model assumes that the county error component is correlated with the regressors. Comparing the two-way pooled OLS and fixed effects models using, we find (as anticipated) that the fixed effects model provides a better model fit. For the fixed effects model, we apply OLS to transformed data, which provides consistent estimates of $\beta$, but not efficient estimates if the error is serially correlated. We find that the error is serially correlated, based on tests for serial correlation. To control for serial correlation, the standard errors are corrected by clustering following the approach in Arellano (1987).

We also fit a fixed effects model using generalized least squares. A comparison of the OLS and GLS fixed effects models (accounting for county fixed effects), as provided in Table 16, suggests that the models are not substantially different; thus, further models use OLS, which allows for inclusion of additional control variables and accounting for both county and year fixed effects.

We additionally fit a first-difference model, which removes time-invariant individual components by lagging the model and subtracting:

$$\Delta Y_{ct} = \beta \Delta ProducingWells_{ct} + \Delta X_{ct}\theta + \Delta \varepsilon_{ct}$$  \hspace{1cm} (60)

The first-difference model is efficient and usually preferred if the error is persistent over time because $\Delta \varepsilon_{ct}$ will be serially uncorrelated. A summary of the coefficients, standard errors, and model fit between the fixed effects and first-difference models are summarized in Table 14.

For the proceeding model specifications, we use OLS fixed effects, including county and year fixed effects, and controlling for serial correlation by clustering standard errors by county.

---

1 We employ the King/Wu and Breusch/Pagan statistics tests for one- and two-way unbalanced panel data as described in Baltagi (2013) and implemented in the plm R package. The following are the test statistics and p-values for each model form, where the null hypothesis is:

| Test Statistic | p-value | Test Statistic | p-value |
|----------------|---------|----------------|---------|
| Two-way        | 28.158  | < 2.2e-16      | 16767   | < 2.2e-16 |
| County effects | 129.49  | < 2.2e-16      | 16767   | < 2.2e-16 |
| Year effects   | 0.36832 | 0.3563         | 0.13566 | 0.7126   |

2 We employ the F-test for two-way effects, used for comparing within and pooling models, and implemented in the plm R package. The following are the test statistic and p-value, where the null hypothesis is that there are no significant fixed effects: $F = 459.54$, df1 = 268, df2 = 2770, p-value < 2.2e-16

3 We employ the Breusch-Godfrey/Wooldridge test for serial correlation in panel models as described in Wooldridge (2010) and implemented in the plm R package. The following is the test statistic, where the null hypothesis is that there is no serial correlation: chisq = 750.25, df = 1, p-value = 6.796e-07

We employ the Wooldridge’s test for serial correlation, which is applicable to fixed effects panels, in particular to short panels with small T and large n. The following is the test statistic, where the null hypothesis is that there is no serial correlation: $F = 24.78$, df1 = 1, df2 = 2952, p-value = 6.796e-07
### Table 13. County-level descriptive statistics for different county subsets over the period 2004 to 2016.

| Variable                                      | Units                        | Pennsylvania, Ohio, West Virginia, and New York (269) | Pennsylvania, Ohio, West Virginia, and New York shale counties (193) | Pennsylvania, Ohio, West Virginia, and New York nonmetropolitan counties (241) | Pennsylvania, Ohio, West Virginia, and New York shale and nonmetropolitan counties (186) |
|-----------------------------------------------|------------------------------|------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
|                                               |                              | Mean        | Std. dev. | Mean        | Std. dev. | Mean        | Std. dev. | Mean        | Std. dev. | Mean        | Std. dev. |
| Employment                                    | jobs                         | 101,979     | 244,794   | 61,785      | 125,765   | 42,303      | 44,139    | 40,591      | 43,052    |
| Income                                        | $'000                        | 7,827,151   | 19,421,640| 4,094,918   | 7,789,432 | 3,002,160   | 3,183,225 | 2,796,412   | 2,865,572 |
| Unconventional production                     | Mmcf                         | 7,727       | 59,887    | 10,951      | 71,048    | 8,599       | 63,303    | 11,330      | 72,455    |
| Unconventional producing wells                | wells                        | 15          | 80        | 21          | 95        | 17          | 85        | 22          | 97        |
| Unconventional spud wells                     | wells                        | 3           | 18        | 5           | 21        | 4           | 19        | 5           | 21        |
| Earnings                                      | $'000                        | 5,852,333   | 22,304,600| 6,904,864   | 2,796,412 | 2,865,572   | 2,197,252 |
| Farm share of earnings                        | $/$                          | 0.01        | 0.02      | 0.01        | 0.02      | 0.01        | 0.02      | 0.01        | 0.02      |
| Construction share of earnings                | $/$                          | 0.07        | 0.03      | 0.07        | 0.04      | 0.07        | 0.03      | 0.07        | 0.04      |
| Manufacturing share of earnings               | $/$                          | 0.16        | 0.11      | 0.15        | 0.10      | 0.17        | 0.11      | 0.16        | 0.10      |
| Retail share of earnings                      | $/$                          | 0.07        | 0.02      | 0.07        | 0.02      | 0.07        | 0.02      | 0.07        | 0.02      |
| Population                                   | persons                      | 176,021     | 335,726   | 109,364     | 180,667   | 82,419      | 74,979    | 78,762      | 70,167    |
| Population density                           | persons per mile             | 971         | 5,584     | 200         | 164       | 195         | 148       | 146         |           |
| Employment rate                              | %                            | 93          | 2         | 93          | 2         | 93          | 2         | 93          | 2         |
| Employment ratio                             | jobs per person              | 0.49        | 0.14      | 0.48        | 0.12      | 0.48        | 0.11      | 0.47        | 0.11      |

### Table 14. Employment effects for pooled OLS, first-differenced, and fixed effects (implemented using OLS).

| Variables                                      | (1) County and year fixed effects (OLS) | (2) Pooled OLS | (3) First-differenced |
|-----------------------------------------------|----------------------------------------|----------------|-----------------------|
|                                               | Estimate        | Std. Error | P>|t|     | Estimate        | Std. Error | P>|t|     | Estimate        | Std. Error | P>|t|     |
| Producing wells                               | 5.36           | 1.86       | 0.004    **           | 2.28           | 2.85       | 0.424    | 5.66           | 1.28       | 0.000    ***    |
| Population                                    | 0.47           | 0.09       | 0.000    ***          | 0.63           | 0.00       | <2e-16    ***    | 0.43           | 0.03       | <2e-16    ***    |
| Population density                            | 88.91          | 31.63      | 0.005    **           | -15.01         | 0.75       | <2e-16    ***    | 106.73         | 7.42       | <2e-16    ***    |
| Farm earnings share                           | 6.88E+03       | 3.30E+03   | 0.037    *            | -1.67E+03      | 9.98E+03   | 0.867    | -1.26E+03    | 2.45E+03   | 0.607    |
| Construction earnings share                   | 8.08E+03       | 3.79E+03   | 0.033    *            | -1.26E+04      | 6.48E+03   | 0.052    | 9.77E+03    | 2.05E+03   | 0.000    ***   |
| Manufacturing earnings share                  | 7.26E+03       | 3.24E+03   | 0.025    *            | 1.05E+02       | 2.01E+03   | 0.958    | 1.57E+04    | 1.86E+03   | 0.000    ***   |
| Retail earnings share                         | -3.77E+04      | 1.25E+04   | 0.003    **           | -8.76E+04      | 7.83E+03   | <2e-16    ***    | -3.13E+04    | 5.96E+03   | 0.000    ***   |
| Number of counties                            | 241            |            |           | 241            |            |           | 241            |            |           |
| Years                                         | 2004-2016      |            |           | 2004-2016      |            |           | 2004-2016      |            |           |
| Sample                                       | Nonmetropolitan counties |            |           | Nonmetropolitan counties |            |           | Nonmetropolitan counties |            |           |
| R-squared                                    | 0.55           |            |           | 0.97           |            |           | 0.28           |            |           |
| Adjusted R-squared                           | 0.51           |            |           | 0.97           |            |           | 0.28           |            |           |

Significance codes: ***p<0.001, **p<0.01, *p<0.05, .p<0.1
Table 15. Fixed effects assuming two-way, county, and time fixed effects.

| Variables                  | (1) County and year fixed effects (OLS) | (4) County fixed effects (OLS) | (5) Year fixed effects (OLS) |
|----------------------------|----------------------------------------|--------------------------------|-----------------------------|
|                            | Estimate  | Std. Error | P>|t| < 0.05 | Estimate  | Std. Error | P>|t| < 0.05 | Estimate  | Std. Error | P>|t| < 0.05 |
| Producing wells            | 5.36      | 1.86       | 0.004  **   | 6.08      | 1.89       | 0.001  **   | -15.76    | 6.64       | 0.018  *   |
| Population                 | 0.47      | 0.09       | 0.000  ***  | 0.46      | 0.09       | 0.000  ***  | 0.42      | 0.04       | < 2e-16    ***|
| Population density         | 88.91     | 31.63      | 0.005  **   | 92.34     | 31.69      | 0.004  **   | 21.36     | 1.19       | < 2e-16    ***|
| Farm earnings share        | 6.88E+03  | 3.30E+03   | 0.037  *    | -8.06E+03 | 3.33E+03   | 0.016  *    | -1.13E+05 | 5.76E+04   | 0.050  *    |
| Construction earnings share| 8.08E+03  | 3.79E+03   | 0.033  *    | 1.46E+04  | 4.26E+03   | 0.001  ***  | -2.09E+05 | 4.37E+04   | 0.000  ***  |
| Manufacturing earnings share| 7.26E+03 | 3.24E+03   | 0.025  *    | 1.41E+04  | 3.86E+03   | 0.000  ***  | -1.06E+05 | 2.51E+04   | 0.000  ***  |
| Retail earnings share      | -3.77E+04 | 1.25E+04   | 0.003  **   | -3.73E+04 | 1.10E+04   | 0.001  ***  | -7.26E+05 | 1.63E+05   | 0.000  ***  |

Number of counties 241
Years 2004-2016
Sample Nonmetropolitan counties
R-squared 0.55
Adjusted R-squared 0.51

Significance codes: ***p<0.001, **p<0.01, *p<0.05, .p<0.1

Table 16. Employment effects for fixed effects implemented using OLS and fixed effects implemented using GLS.

| Variables                  | (4) County fixed effects (OLS) | (6) County fixed effects (GLS) |
|----------------------------|--------------------------------|--------------------------------|
|                            | Estimate  | Std. Error | P>|t| < 0.05 | Estimate  | Std. Error | P>|t| < 0.05 |
| Producing wells            | 6.08      | 1.89       | 0.001  **   | 5.66E+00  | 1.28       | 0.000  *** |
| Population                 | 0.46      | 0.09       | 0.000  ***  | 4.28E-01  | 0.03       | < 2e-16    ***|
| Population density         | 92.34     | 31.69      | 0.004  **   | 1.07E+02  | 7.42E+00   | < 2e-16    ***|
| Farm earnings share        | -8.06E+03 | 3.33E+03   | 0.016  *    | -1.26E+03 | 2446.80    | 0.607     |
| Construction earnings share| 1.46E+04  | 4.26E+03   | 0.001  ***  | 9.77E+03  | 2048.70    | 0.000     |
| Manufacturing earnings share| 1.41E+04 | 3.86E+03   | 0.000  ***  | 1.57E+04  | 1856.10    | < 2e-16    ***|
| Retail earnings share      | -3.73E+04 | 1.10E+04   | 0.001  ***  | -3.13E+04 | 5963.40    | 0.000     |

Number of counties 241
Years 2004-2016
Sample Nonmetropolitan counties
R-squared 0.54
Adjusted R-squared 0.49

Significance codes: ***p<0.001, **p<0.01, *p<0.05, .p<0.1

42
Data subsets

We specify fixed effects models for various subsets of the full panel dataset, partitioned based on the classification of counties as nonmetropolitan and/or shale. Several other empirical studies, including Parades (2015), partition or exclude counties which are not within the Marcellus (or respective shale formation). Including only nonmetropolitan counties creates a more homogenous sample, precluding counties with large cities from excessively influencing estimates from a linear model (Weber 2014). Moreover, nonmetropolitan counties with thin labor markets are presumably the population of interest.

We find that the effect sizes on the producing well variable are fairly consistent across models, as shown in Table 17. The model fit using the shale county data subset has an overall model fit similar to that for model based on the full panel dataset, with a lower standard error for the producing well coefficient. The other models fit using data subsets do not provide as good of a fit as the model based on the full panel dataset.

Table 17. Employment effects for different subsets of data.

| Variables                      | (1) Nonmetropolitan counties | (7) All counties | (8) Shale counties | (9) Shale and nonmetropolitan counties |
|-------------------------------|------------------------------|------------------|-------------------|---------------------------------------|
|                               | Estimate         | Std. Error   | P>|t|    | Estimate        | Std. Error   | P>|t|    | Estimate        | Std. Error   | P>|t|    | Estimate        | Std. Error   | P>|t|    |
| Producing wells               | 5.36            | 1.86        | 0.004 ** | 3.52             | 2.14        | 0.101 | 4.75             | 1.97        | 0.016 * | 5.93             | 1.83        | 0.001 ** |
| Population                    | 0.47            | 0.09        | 0.000 ***| 0.47             | 0.16        | 0.003 **| 0.87             | 0.44        | 0.046 * | 0.64             | 0.17        | 0.000 ***|
| Population density            | 88.91           | 31.63       | 0.005 *  | 74.80           | 21.10       | 0.000 ***| -179.61         | 232.38      | 0.440   | -43.13           | 85.39       | 0.614   |
| Farm earnings share           | 6.88E+03        | 3.30E+03    | 0.037   | 2.70E+04        | 9.71E+03    | 0.005 **| 3.55E+03        | 4.56E+03    | 0.436   | -1.24E+02        | 2.96E+03    | 0.967   |
| Construction earnings share   | 8.08E+03        | 3.79E+03    | 0.033   | 1.37E+04        | 6.89E+03    | 0.047   | 4.34E+03        | 5.10E+03    | 0.395   | 5.28E+03         | 2.60E+03    | 0.042   |
| Manufacturing earnings share  | 7.26E+03        | 3.24E+03    | 0.025   | 9.06E+03        | 4.78E+03    | 0.058   | 8.99E+03        | 3.74E+03    | 0.016   | 9.38E+03         | 3.38E+03    | 0.006   **|
| Retail earnings share         | -3.77E+04       | 1.25E+04    | 0.003   | -3.37E+04       | 3.19E+04    | 0.291   | -2.37E+04       | 1.62E+04    | 0.143   | -1.44E+04        | 8.62E+03    | 0.096   |
| Number of counties            | 241             |             |        | 269             |             |        | 193             |             |        | 186             |             |        |
| Years                         | 2004-2016       |             |        | 2004-2016       |             |        | 2004-2016       |             |        | 2004-2016       |             |        |
| Sample                        | Nonmetropolitan counties |             |        | All counties    |             |        | Shale counties  |             |        | Shale and nonmetropolitan counties |             |        |
| R-squared                     | 0.55            |             |        | 0.76            |             |        | 0.45            |             |        | 0.56            |             |        |
| Adjusted R-squared            | 0.51            |             |        | 0.73            |             |        | 0.40            |             |        | 0.51            |             |        |

Significance codes: ***p<0.001, **p<0.01, *p<0.05, .p<0.1
Variable selection

We specify various fixed effects models with alternative variables, including i) lag and lead variables to capture the dynamic effect of natural gas development, ii) natural gas activity variables (i.e., producing wells, production, and spud wells), iii) natural gas development and time interactions, and iv) employment-to-population ratio and employment rate (as alternative dependent variables).

The following specification includes lag and lead variables to assess the dynamic effect of natural gas development:

$$Y_{ct} = \beta \text{ProducingWell}_{ct-t} + X_{ct} \theta + \alpha_c + \gamma_t + \epsilon_{ct}$$ (61)

Including lag and lead variables allows the estimation of long-run effects which may result from support industries entering the area. Given that unconventional natural gas activity is highly correlated over time, including lags and leads of producing wells induces multicollinearity and makes inference more challenging (Paredes 2015). As shown in Table 18, variants of the distributed lag model show that the aggregated effect of the contemporaneous and lag/lead variables yield a similar effect size as that for the contemporaneous specification. In addition, the lag and lead variables are statistically significant.

We also fit models for different natural gas activity variables, including producing wells, production, and spud wells. As shown in Table 20, the effects associated with alternatively incorporating a single type of activity variable are all positive and significant. When including only the producing wells variable, we find the mean marginal effect size is 5 jobs per producing well; Paredes (2015) similarly finds that 6.8 to 16.8 jobs are supported per producing well. Given that producing wells and production are highly correlated, including both terms is redundant. We find that including both producing and spud wells reveals positive and significant effects associated with both variables. This can be interpreted as a larger employment effect associated with well development than that associated with an already producing well.

Incorporating an interaction term between gas activity and a dummy variable indicating whether a well was producing before or after 2012, we find decreasing marginal employment effects from natural gas activity over time, with 16 jobs per producing well prior to 2012 and 5 jobs thereafter. The intuition is that learning occurs and the industry becomes more efficient over time, which translates into decreasing marginal employment effects from natural gas activity over time. Table 22 shows the interaction between natural gas activity and time.

Finally, we specify models with alternative dependent variables, including the employment-to-population ratio and employment rate. The purpose of using rate-based dependent variables is because they may have a more meaningful interpretation than absolute employment; for example, the creation of 100 jobs in a given county may or may not be significant, depending on the size of the total labor market. As previously noted, the underlying data used to develop the employment rate may not accurately reflect annual changes disaggregated at the county-level, thus limiting their utility for regression analysis. We find that the models incorporating these alternative dependent variables do not provide a good overall model fit; however, there is a positive and significant effect on the producing well variable for the employment-to-population ratio model.
Table 18. Employment effects for distributed lag compared to contemporaneous models.

| Variables               | (1) No lag          | (10) 1-year lag         | (11) 2-year lag        |
|------------------------|---------------------|-------------------------|------------------------|
|                        | Estimate | Std. Error | P>|t| | Estimate | Std. Error | P>|t| | Estimate | Std. Error | P>|t| |
| Producing wells        | 5.36     | 1.86       | 0.004 **               | 19.64     | 5.45       | 0.000 ***  | 11.13     | 3.58       | 0.002 **   |
| 1-year lag producing wells | -16.68 | 4.60       | 0.000 ***                  | -9.18     | 3.21       | 0.004 ** |
| 2-year lag producing wells | -       | -          | -                       | -         | -          | -          |
| Population             | 0.47     | 0.09       | 0.000 ***               | 0.49      | 0.15       | 0.001 ***  | 0.52      | 0.15       | 0.001 ***  |
| Population density     | 88.91    | 31.63      | 0.005 **               | 72.64     | 70.43      | 0.302      | 61.02     | 76.14      | 0.423      |
| Farm earnings share    | 6.88E+03 | 3.30E+03   | 0.037 **               | 7.91E+03  | 3.69E+03   | 0.032 *     | 8.99E+03  | 4.07E+03  | 0.027 *     |
| Construction earnings share | 8.08E+03 | 3.79E+03 | 0.033 * | 5.86E+03 | 3.91E+03 | 0.134 | 4.87E+03 | 3.87E+03 | 0.208 |
| Manufacturing earnings share | 7.26E+03 | 3.24E+03 | 0.025 | 6.76E+03 | 3.48E+03 | 0.052 | 5.66E+03 | 3.78E+03 | 0.135 |
| Retail earnings share  | -3.77E+04 | 1.25E+04 | 0.003 ** | -3.30E+04 | 1.44E+04 | 0.022 | -3.37E+04 | 1.69E+04 | 0.046 |
| Number of counties     | 241      | 241        | 241                     |
| Years                  | 2004-2016 | 2004-2016  | 2004-2016               |
| Sample                 | Nonmetropolitan counties | Nonmetropolitan counties | Nonmetropolitan counties |
| R-squared              | 0.51     | 0.50       | 0.46                    |
| Adjusted R-squared     | 0.51     | 0.45       | 0.39                    |

Significance codes: ***p<0.001, **p<0.01, *p<0.05, .p<0.1

Table 19. Employment effects for distributed lead compared to contemporaneous models.

| Variables               | (1) No lag          | (12) 1-year lead         | (13) 2-year lead        |
|------------------------|---------------------|-------------------------|------------------------|
|                        | Estimate | Std. Error | P>|t| | Estimate | Std. Error | P>|t| | Estimate | Std. Error | P>|t| |
| Producing wells        | 5.36     | 1.86       | 0.004 **               | -12.27    | 4.04       | 0.002 **  | -3.31     | 2.69       | 0.220      |
| 1-year lead producing wells | 17.39 | 5.12       | 0.001 ***               | -         | -          | -          | 9.46      | 3.53       | 0.007 **   |
| 2-year lead producing wells | -       | -          | -                       | -         | -          | -          |
| Population             | 0.47     | 0.09       | 0.000 ***               | 0.46      | 0.16       | 0.004 **  | 0.40      | 0.18       | 0.027      |
| Population density     | 88.91    | 31.63      | 0.005 **               | 79.58     | 70.35      | 0.258      | 94.17     | 78.55      | 0.231      |
| Farm earnings share    | 6.88E+03 | 3.30E+03   | 0.037 *               | 7.25E+03  | 4.17E+03   | 0.082      | 6.22E+03  | 5.46E+03   | 0.255      |
| Construction earnings share | 8.08E+03 | 3.79E+03 | 0.033 | 5.11E+03 | 4.51E+03 | 0.257 | 5.83E+03 | 5.58E+03 | 0.297 |
| Manufacturing earnings share | 7.26E+03 | 3.24E+03 | 0.025 | 8.61E+03 | 4.04E+03 | 0.033 | 9.30E+03 | 4.97E+03 | 0.062 |
| Retail earnings share  | -3.77E+04 | 1.25E+04 | 0.003 ** | -3.18E+04 | 1.38E+04 | 0.021 | -3.30E+04 | 1.61E+04 | 0.041 |
| Number of counties     | 241      | 241        | 241                     |
| Years                  | 2004-2016 | 2004-2016  | 2004-2016               |
| Sample                 | Nonmetropolitan counties | Nonmetropolitan counties | Nonmetropolitan counties |
| R-squared              | 0.55     | 0.54       | 0.51                    |
| Adjusted R-squared     | 0.51     | 0.49       | 0.45                    |

Significance codes: ***p<0.001, **p<0.01, *p<0.05, .p<0.1
Table 20. Employment effects for different natural gas activity variables.

| Variables                          | (16) Producing and spud wells | (17) Producing wells and production | (18) Producing wells, spud wells, and production |
|-----------------------------------|-------------------------------|-----------------------------------|--------------------------------------------------|
| Producing wells                   | Estimate                      | Std. Error                        | P>|t|                  | Estimate | Std. Error | P>|t|                  | Estimate | Std. Error | P>|t|                  |
|                                   | 4.50                          | 1.59                              | 0.005 **             | -8.30E-03 | 0.00       | 0.063 *          | 1.07E+01 | 3.74       | 0.004 **            |
| Spud wells                         | 1.64E+01                      | 4.20                              | 0.000 ***            | -                       | -1.14E+01 | 4.70       | 0.015 **          | 8.49E-03  | 0.00        | 0.17 *             |
| Production                         | -                             | -                                 | -                   | -                       | 8.90E+01 | 31.61      | 0.005             | 8.95E+01  | 31.56       | 0.005 **            |
| Population                         | 4.68E-01                      | 8.67E-02                          | 0.000 ***            | 4.70E-01 | 8.66E-02  | 0.000 **          | 4.69E-01 | 8.65E-02  | 0.000 ***            |
| Population density                 | 8.94E+01                      | 31.58                              | 0.000              | 8.90E+01 | 31.61      | 0.005             | 8.95E+01 | 31.56       | 0.005 **            |
| Farm earnings share                | 8.70E+03                      | 3374.10                           | 0.000 **            | 7.17E+03 | 3335.50   | 0.032             | 9.01E+03 | 3389.60   | 0.008 **            |
| Construction earnings share        | 5.75E+03                      | 3.98E+03                          | 0.149               | 7.58E+03 | 3.87E+03  | 0.050             | 5.21E+03 | 4.04E+03  | 0.198              |
| Manufacturing earnings share       | 7.52E+03                      | 3280.50                           | 0.022               | 7.28E+03 | 3226.20   | 0.024             | 7.54E+03 | 3264.50   | 0.021              |
| Retail earnings share              | -1.53E+01                     | 12598.00                          | 0.005 **            | -3.80E+04 | 12516.00  | 0.002             | -3.56E+04 | 12612.00  | 0.005 **            |
| Number of counties                | 241                           |                                   |                     | 241                    | 241             |                     | 241             |                     |
| Years                             | 2004-2016                     |                                   |                     | 2004-2016             | 2004-2016       |                     | 2004-2016       |                     |
| Sample                            | Nonmetropolitan counties      |                                   |                     | Nonmetropolitan counties | Nonmetropolitan counties |                     | Nonmetropolitan counties |                     |
| R-squared                         | 0.56                          |                                   |                     | 0.56                   | 0.56            |                     | 0.56            |                     |
| Adjusted R-squared                | 0.52                          |                                   |                     | 0.52                   | 0.52            |                     | 0.52            |                     |

Table 21. Employment effects for different natural gas activity variables.

| Variables                          | (1) Producing wells | (14) Spud wells | (15) Production |
|-----------------------------------|--------------------|----------------|-----------------|
| Producing wells                   | Estimate | Std. Error | P>|t|      | Estimate | Std. Error | P>|t|      | Estimate | Std. Error | P>|t|      |
|                                   | 5.36     | 1.86       | 0.004 **      | -         | 2.08E+01 | 6.32       | 0.001 **      | -         | 5.01E+03 | 0.00       | 0.037 *      |
| Spud wells                         | -        | -          | -            | -         | -        | -          | -              | -         | -        | -          | -            |
| Production                         | 0.47     | 0.09       | 0.000 **      | 4.63E-01 | 1.85E+05 | 0.000 **      | 4.66E-01 | 8.68E-02 | 0.000 **      |
| Population                         | 88.91    | 31.63      | 0.005 **      | 8.92E+01 | 21.51    | 0.005 **      | 8.87E+01 | 31.63    | 0.005 **      |
| Farm earnings share                | 6.88E+03 | 3.30E+03  | 0.037 *      | 7.35E+03 | 3.43E+03 | 0.032 *      | 8.95E+03 | 3.31E+03 | 0.071 *      |
| Construction earnings share        | 8.08E+03 | 3.79E+03  | 0.033 *      | 7.61E+03 | 4.02E+03 | 0.058 *      | 9.38E+03 | 3.81E+03 | 0.014 *      |
| Manufacturing earnings share       | 7.26E+03 | 3.24E+03  | 0.025 *      | 6.44E+03 | 3.17E+03 | 0.043 *      | 6.79E+03 | 3.20E+03 | 0.034 *      |
| Retail earnings share              | -3.77E+04| 1.25E+04  | 0.003 **      | -4.02E+04| 1.23E+04 | 0.001 **      | -3.97E+04| 1.23E+04 | 0.001 **      |
| Number of counties                | 241      |            |              | 241      |            |              | 241            |            |              |
| Years                             | 2004-2016 |            |              | 2004-2016 |            |              | 2004-2016       |            |              |
| Sample                            | Nonmetropolitan counties |            |              | Nonmetropolitan counties |            |              | Nonmetropolitan counties |            |              |
| R-squared                         | 0.55     |            |              | 0.55     |            |              | 0.55            |            |              |
| Adjusted R-squared                | 0.51     |            |              | 0.51     |            |              | 0.50            |            |              |

Significance codes: ***p<0.001, **p<0.01, *p<0.05, .p<0.1
Table 22. Employment effects including time and natural gas activity interactions.

| Variables                      | Estimate | Std. Error | P>|t|) | Estimate | Std. Error | P>|t| |
|-------------------------------|----------|------------|------|----------|------------|------|
| Producing wells               | 5.36     | 1.86       | 0.004 | **       | 5.62       | 1.91 | 0.003 | ** |
| Population                    | 0.47     | 0.09       | 0.000 | ***      | 4.69E-01   | 0.09 | 0.000 | ***|
| Population density            | 88.91    | 31.63      | 0.005 | **       | 8.90E+01   | 1.36E+01 | 0.005 | ** |
| Farm earnings share           | 6.88E+03 | 3.30E+03   | 0.037 | *        | 7.83E+03   | 3321.50 | 0.018 | * |
| Construction earnings share   | 8.08E+03 | 3.79E+03   | 0.033 | *        | 7.49E+03   | 3709.40 | 0.043 | * |
| Manufacturing earnings share  | 7.26E+03 | 3.24E+03   | 0.025 | *        | 7.21E+03   | 3228.60 | 0.026 | * |
| Retail earnings share         | -3.77E+04| 1.25E+04   | 0.003 | **       | -3.64E+04  | 12459.00 | 0.004 | ** |
| Producing wells x Before 2012 | 1.08E+01 | 1.62       | 0.000 | ***      | 1.08E+01   | 1.62 | 0.000 | ***|
| Number of counties            | 241      |            |      |          | 241        |      |      |    |
| Years                         | 2004-2016|            |      |          | 2004-2016  |      |      |    |
|Sample                         | Nonmetropolitan counties |              |      |          | Nonmetropolitan counties |      |      |    |
|R-squared                      | 0.55     |            |      |          | 0.56       |      |      |    |
|Adjusted R-squared             | 0.51     |            |      |          | 0.51       |      |      |    |

Significance codes: ***p<0.001, **p<0.01, *p<0.05, .p<0.1

Table 23. Effects for models with alternative dependent variables.

| Variables                      | (1) Employment | (20) Employment rate | (21) Employment-to-population ratio |
|-------------------------------|----------------|----------------------|-------------------------------------|
|                                | Estimate | Std. Error | P>|t|) | Estimate | Std. Error | P>|t|) | Estimate | Std. Error | P>|t|) |
| Producing wells               | 5.36     | 1.86       | 0.004 | **       | -1.55E-03 | 0.00 | 0.000 | *** | 5.79E-05 | 0.00 | 0.000 | *** |
| Population                    | 0.47     | 0.09       | 0.000 | ***      | -1.20E-05 | 8.31E-06 | 0.148 | -2.40E-07 | 1.85E-07 | 0.193 |
| Population density            | 88.91    | 31.63      | 0.005 | **       | -2.89E-03 | 0.00 | 0.124 | 2.08E-04 | 0.00 | 0.000 | *** |
| Farm earnings share           | 6.88E+03 | 3.30E+03   | 0.037 | *        | -1.13E+01 | 2.43 | 0.000 | *** | -1.50E-01 | 0.05 | 0.004 | ** |
| Construction earnings share   | 8.08E+03 | 3.79E+03   | 0.033 | *        | 7.27E+00  | 2.31 | 0.002 | **  | 1.28E-01 | 4.73E-02 | 0.007 | ** |
| Manufacturing earnings share  | 7.26E+03 | 3.24E+03   | 0.025 | *        | 2.18E+00  | 2.21 | 0.324 | 8.81E-02 | 0.05 | 0.058 | . |
| Retail earnings share         | -3.77E+04| 1.25E+04   | 0.003 | **       | -9.50E+00 | 6.37 | 0.136 | -6.91E-01 | 0.21 | 0.001 | *** |
| Number of counties            | 241      |            |      |          | 241       |      |      |    |
| Years                         | 2004-2016|            |      |          | 2004-2016 |      |      |    |
|Sample                         | Nonmetropolitan counties |              |      |          | Nonmetropolitan counties |      |      |    |
|R-squared                      | 0.55     |            |      |          | 0.04      |      |      |    |
|Adjusted R-squared             | 0.51     |            |      |          | -0.05     |      |      |    |

Significance codes: ***p<0.001, **p<0.01, *p<0.05, .p<0.1
Model checking

Actual versus predicted employment

We compare the actual versus predicted employment for some of the models specified in the previous sections. All of the model predictions closely fit the actual employment.

Figure 25. Actual versus predicted employment for different model specifications.

Bootstrapping

To capture the uncertainty around the employment effects of natural gas activity, we estimate confidence intervals through bootstrapping. We use block sampling, where a single draw consists of all observations for a county, thereby maintaining the dependence structure between years. We also resample cases, rather than residuals, which does not assume that the shape of regression function or the distribution of the error of the original model are correct; resampling cases treats each bootstrap sample as an observation and refits the fixed effects model.
Figure 26. Annual employment based on model specifications 1, 14, and 16. Solid line represents mean. Shaded region represents 95% confidence interval based on robust standard errors clustered by county. Dashed lines represent bootstrapped 95% confidence interval. Dotted lines represent within model 95% confidence interval.
Supplementary References

1. U.S. Geological Survey. National Assessment of Oil and Gas Project - Appalachian Basin Province (067) Boundary. (2002).
2. U.S. Geological Survey. National Assessment of Oil and Gas Project Devonian Marcellus Shale of the Appalachian Basin Province (067) Assessment Units. (2011).
3. U.S. Geological Survey. National Assessment of Oil and Gas Project - Utica Shale Unconventional Assessment. (2015).
4. Pennsylvania Department of Environmental Protection. PA DEP Oil & Gas Reporting Website. Available at: https://www.paoilandgasreporting.state.pa.us/publicreports/Modules/Welcome/Welcome.aspx.
5. Ohio Department of Natural Resources. Oil & Gas Resources. Available at: http://oilandgas.ohiodnr.gov/production.
6. West Virginia Department of Environmental Protection. Oil and Gas Production Data. Available at: https://dep.wv.gov/oil-and-gas/databaseinfo/Pages/default.aspx.
7. U.S. Energy Information Administration. Natural Gas Gross Withdrawals from Gas Wells. Available at: http://www.eia.gov/dnav/ng/ng_prod_sum_a_epg0_fgg_mmcf_a.htm%
8. U.S. Energy Information Administration. International and Interstate Movements of Natural Gas by State. Available at: http://www.eia.gov/dnav/ng/ng_move_ist_a2dcu_swv_a.htm%
9. U.S. Energy Information Administration. Natural Gas Processed. Available at: http://www.eia.gov/dnav/ng/ng_prod_pp_a_epg0_ygp_mmcf_a.htm%
10. U.S. Energy Information Administration. Natural Gas Lease Fuel Consumption. Available at: http://www.eia.gov/dnav/ng/ng_cons_sum_a_epg0_vcl_mmcf_a.htm%
11. U.S. Energy Information Administration. Natural Gas Plant Fuel Consumption. Available at: http://www.eia.gov/dnav/ng/ng_cons_sum_a_epg0_vcf_mmcf_a.htm%
12. U.S. Energy Information Administration. Natural Gas Pipeline & Distribution Use. Available at: http://www.eia.gov/dnav/ng/ng_cons_sum_a_epg0_vgp_mmcf_a.htm%
13. U.S. Energy Information Administration. Natural Gas Delivered to Electric Power Consumers. Available at: http://www.eia.gov/dnav/ng/ng_cons_sum_a_epg0_veu_mmcf_a.htm%
14. U.S. Energy Information Administration. Natural Gas Delivered to Industrial Consumers. Available at: http://www.eia.gov/dnav/ng/ng_cons_sum_a_epg0_vin_mmcf_a.htm%
15. U.S. Energy Information Administration. Natural Gas Delivered to Commercial Consumers. Available at: http://www.eia.gov/dnav/ng/ng_cons_sum_a_epg0_vcs_mmcf_a.htm%
16. U.S. Energy Information Administration. Natural Gas Delivered to Vehicle Fuel Consumers. Available at: http://www.eia.gov/dnav/ng/ng_cons_sum_a_epg0_vdv_mmcf_a.htm%
17. U.S. Energy Information Administration. Natural Gas Delivered to Residential Consumers. Available at: http://www.eia.gov/dnav/ng/ng_cons_sum_a_epg0_vrs_mmcf_a.htm%
18. Allen, D. T. Atmospheric Emissions and Air Quality Impacts from Natural Gas Production and Use. *Annu. Rev. Chem. Biomol. Eng.* 5, 55–75 (2014).
19. U.S. Environmental Protection Agency. Part II AGENCY: Oil and Natural Gas Sector: New Source Performance Standards and National Emission Standards for Hazardous Air Pollutants Reviews; Final Rule. 77, 1–27 (2012).
20. Roy, A. A., Adams, P. J. & Robinson, A. L. Air pollutant emissions from the development,
production, and processing of Marcellus Shale natural gas. *J. Air Waste Manage. Assoc.* **64**, 19–37 (2014).

21. Roohani, Y. H., Roy, A. A., Heo, J., Robinson, A. L. & Adams, P. J. Impact of natural gas development in the Marcellus and Utica shales on regional ozone and fine particulate matter levels. *Atmos. Environ.* **155**, 11–20 (2017).

22. Jiang, M., Hendrickson, C. T. & Vanbiesen, J. M. Life cycle water consumption and wastewater generation impacts of a Marcellus shale gas well. *Environ. Sci. Technol.* **48**, 1911–1920 (2014).

23. Rahm, B. G. *et al.* Wastewater management and Marcellus Shale gas development: Trends, drivers, and planning implications. *J. Environ. Manage.* **120**, 105–113 (2013).

24. Bar-Ilan, A., Parikh, R., Grant, J., Shah, T. & Pollack, A. K. Recommendations for Improvements to the CENRAP States’ Oil and Gas Emissions Inventories. (2008).

25. Roy, A. A. *et al.* Air pollutant emissions from the development, production, and processing of Marcellus Shale natural gas. *J. Air Waste Manage. Assoc.* **64**, 19–37 (2014).

26. Allen, D. T. *et al.* Methane emissions from process equipment at natural gas production sites in the United States: liquid unloadings. *Env. Sci Technol* **49**, 641–648 (2014).

27. Allen, D. T. *et al.* Measurements of methane emissions at natural gas production sites in the United States. *Proc. Natl. Acad. Sci.* **110**, 17768–17773 (2013).

28. Omara, M. *et al.* Methane Emissions from Conventional and Unconventional Natural Gas Production Sites in the Marcellus Shale Basin. *Environ. Sci. Technol.* **50**, 2099–2107 (2016).

29. U.S. Environmental Protection Agency. Median Life, Annual Activity, and Load Factor Values for Nonroad Engine Emissions Modeling. EPA-420-R-10-016. NR-005 d. U.S. Environ. Agency 1–47 (2010).

30. U.S. Environmental Protection Agency. Nonroad Engine Population Estimates Nonroad Engine Population Estimates. (2010).

31. U.S. Environmental Protection Agency. Calculation of Age Distributions in the Nonroad Model: Growth and Scrappage. (2005).

32. Baker, R. & Pring, M. Drilling Rig Emission Inventory for the State of Texas. 2009, (2014).

33. Armendariz, A., Ph, D. & Alvarez, R. Emissions from Natural Gas Production in the Barnett Shale Area and Opportunities for Cost-Effective Improvements report by: 45 (2009). doi:10.1656/058.012.s413

34. U.S. Environmental Protection Agency. Control Techniques Guidelines for the Oil and Natural Gas Industry. 343 (2016).

35. U.S. Environmental Protection Agency. National Emissions Inventory. (2018). Available at: https://www.epa.gov/air-emissions-inventories/2014-national-emissions-inventory-nei-data.

36. U.S. Environmental Protection Agency. Continuous Environmental Monitoring System. Available at: https://ampd.epa.gov/ampd/.

37. U.S. Environmental Protection Agency. Emissions & Generation Resource Integrated Database. (2018). Available at: https://www.epa.gov/energy/emissions-generation-resource-integrated-database-egrid.

38. Behrer, A. P. & Mauter, M. S. Allocating Damage Compensation in a Federalist System: Lessons from Spatially Resolved Air Emissions in the Marcellus. *Environ. Sci. Technol.* **51**, 3600–3608 (2017).
39. Lutz, B. D., Lewis, A. N. & Doyle, M. W. Generation, transport, and disposal of wastewater associated with Marcellus Shale gas development. *Environ. Res. Lett.* **6**, 647–656 (2011).

40. Jiang, M. *et al.* Life cycle greenhouse gas emissions of Marcellus shale gas. *Environ. Res. Lett.* **6**, 49, 647–656 (2013).

41. Tong, F., Jaramillo, P. & Azevedo, I. M. L. Comparison of Life Cycle Greenhouse Gases from Natural Gas Pathways for Medium and Heavy-Duty Vehicles. *Environ. Sci. Technol.* **49**, 7123–7133 (2015).

42. Weber, C. L. & Clavin, C. Life Cycle Carbon Footprint of Shale Gas: Review of Evidence and Implications. *Environ. Sci. Technol.* **46**, 5688–5695 (2012).

43. U.S. Environmental Protection Agency. *2014 National Emissions Inventory, version 2 - Technical Support Document.* (2018).

44. U.S. Environmental Protection Agency. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2016.* (2018).

45. Venkatesh, A., Jaramillo, P., Griffin, W. M. & Matthews, H. S. Uncertainty in life cycle greenhouse gas emissions from United States natural gas end-uses and its effects on policy. *Environ. Sci. Technol.* **45**, 8182–8189 (2011).

46. U.S. Environmental Protection Agency. *Oil and Natural Gas Sector: Emission Standards for New, Reconstructed, and Modified Sources.* **81**, 35824–35942 (2016).

47. Zimmerle, D. J. *et al.* Methane Emissions from the Natural Gas Transmission and Storage System in the United States. *Environ. Sci. Technol.* **49**, 9374–9383 (2015).

48. Marchese, A. J. *et al.* Methane emissions from United States natural gas gathering and processing-SI. *Environ. Sci. Technol.* (2015).

49. Lamb, B. K. *et al.* Direct and Indirect Measurements and Modeling of Methane Emissions in Indianapolis, Indiana. *Environ. Sci. Technol.* **50**, 8910–8917 (2016).

50. Lamb, B. K. *et al.* Direct measurements show decreasing methane emissions from natural gas local distribution systems in the United States. *Environ. Sci. Technol.* **49**, 5161–9 (2015).

51. Mitchell, A. L. *et al.* Measurements of Methane Emissions from Natural Gas Gathering Facilities and Processing Plants: Measurement Results. *Environ. Sci. Technol.* **49**, 3219–3227 (2015).

52. Myhre, G. *et al.* Anthropogenic and Natural Radiative Forcing. *Clim. Chang. 2013 Phys. Sci. Basis. Contrib. Work. Gr. I to Fifth Assess. Rep. Intergov. Panel Clim. Chang.* 659–740 (2013). doi:10.1017/ CBO9781107415324.018

53. Alvarez, R. A. *et al.* Assessment of methane emissions from the U.S. oil and gas supply chain. *7204*, 1–9 (2018).

54. U.S. Environmental Protection Agency. Annexes to the Inventory of U.S. GHG Emissions and Sinks. 1–458 (2018).

55. U.S. Environmental Protection Agency. *AP 42, Fifth Edition, Volume I Chapter 1: External Combustion Sources.* (1998).

56. U.S. Energy Information Adminstration. Carbon Dioxide Emissions Coefficients. (2016). Available at: https://www.eia.gov/environment/emissions/co2_vol_mass.php.

57. U.S. Energy Information Adminstration. Heat Content of Natural Gas Delivered to Consumers. Available at: http://www.eia.gov/dnav/ng/ng_cons_heat_a_epg0_ygth_btucf_a.htm%0A.

58. Marchese, A. J. *et al.* Methane Emissions from United States Natural Gas Gathering and Processing. *Environ. Sci. Technol.* **49**, 10718–10727 (2015).
59. Pipeline and Hazardous Materials Safety Administration. Gas Distribution, Gas Gathering, Gas Transmission, Hazardous Liquids, Liquefied Natural Gas (LNG), and Underground Natural Gas Storage (UNGS) Annual Report Data. Available at: https://www.phmsa.dot.gov/data-and-statistics/pipeline/gas-distribution-gas-gathering-gas-transmission-hazardous-liquids.

60. U.S. Energy Information Administration. Natural Gas Delivered to Consumers. Available at: http://www.eia.gov/dnav/ng/ng_cons_sum_a_epg0_vgt_mmcf_a.htm.

61. Frey, H. C. & Zhao, Y. Quantification of variability and uncertainty for air toxic emission inventories with censored emission factor data. *Environ. Sci. Technol.* **38**, 6094–6100 (2004).

62. Brandt, A. R., Heath, G. A. & Cooley, D. Methane Leaks from Natural Gas Systems Follow Extreme Distributions. *Environ. Sci. Technol.* **50**, 12512–12520 (2016).

63. Romano, D., Bernetti, A. & De Lauretis, R. Different methodologies to quantify uncertainties of air emissions. *Environ. Int.* **30**, 1099–1107 (2004).

64. Burnham, A. *et al.* Life-cycle greenhouse gas emissions of shale gas, natural gas, coal, and petroleum. *Environ. Sci. Technol.* **46**, 619–627 (2012).

65. Zimmerle, D. J. *et al.* Gathering pipeline methane emissions in Fayetteville shale pipelines and scoping guidelines for future pipeline measurement campaigns. *Elem Sci Anth.* **5**, 70 (2017).

66. Shine, K. P., Fuglestvedt, J. S., Hailemariam, K. & Stuber, N. Alternatives to the Global Warming Potential for comparing climate impacts of emissions of greenhouse gases. *Clim. Change* **68**, 281–302 (2005).

67. Boucher, O. & Reddy, M. S. Climate trade-off between black carbon and carbon dioxide emissions. *Energy Policy* **36**, 193–200 (2008).

68. Shine, K. P., Berntsen, T. K., Fuglestvedt, J. S., Skeie, R. B. & Stuber, N. Comparing the climate effect of emissions of short- and long-lived climate agents. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **365**, 1903–1914 (2007).

69. Tol, R. S. J., Berntsen, T. K., Oneill, B. C., Fuglestvedt, J. S. & Shine, K. P. A unifying framework for metrics for aggregating the climate effect of different emissions. *Environ. Res. Lett.* **7**, (2012).

70. Shindell, D. *et al.* Climate, health, agricultural and economic impacts of tighter vehicle-emission standards. *Nat. Clim. Chang.* **1**, 59–66 (2011).

71. Berntsen, T. & Fuglestvedt, J. Global temperature responses to current emissions from the transport sectors. *Proc. Natl. Acad. Sci.* **105**, 19154–19159 (2008).

72. Peters, G. P., Aamaas, B., Berntsen, T. & Fuglestvedt, J. S. The integrated global temperature change potential (iGTP) and relationships between emission metrics. *Environ. Res. Lett.* **6**, 044021 (2011).

73. Fuglestvedt, J. S. *et al.* Transport impacts on atmosphere and climate: Metrics. *Atmos. Environ.* **44**, 4648–4677 (2010).

74. Joos, F. *et al.* Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: A multi-model analysis. *Atmos. Chem. Phys.* **13**, 2793–2825 (2013).

75. Collins, W. J. *et al.* Global and regional temperature-change potentials for near-term climate forcers. *Atmos. Chem. Phys.* **13**, 2471–2485 (2013).

76. National Oceanic and Atmospheric Administration. Global atmospheric carbon dioxide concentration. Available at: ftp://aftp.cmdl.noaa.gov/products/trends/co2/co2_annmean_mlo.txt.

77. Boucher, O. Comparison of physically- and economically-based CO2- equivalences for methane. *Earth Syst. Dyn.* **3**, 49–61 (2012).
78. Allen, M. R. et al. New use of global warming potentials to compare cumulative and short-lived climate pollutants. *6*, (2016).

79. Smith, S. M. et al. Equivalence of greenhouse-gas emissions for peak temperature limits. *Nat. Clim. Chang.* 2,1–4 (2012).

80. Shoemaker, J. K., Schrag, D. P., Molina, M. J. & Ramanathan, V. What Role for Short-Lived Climate Pollutants in Mitigation Policy? *Science (80-. ).* 342, 1323–1325 (2013).

81. Balcombe, P., Anderson, K., Speirs, J., Brandon, N. & Hawkes, A. The Natural Gas Supply Chain: The Importance of Methane and Carbon Dioxide Emissions. *ACS Sustain. Chem. Eng.* 5, 3–20 (2017).

82. U.S. Bureau of Economic Analysis. Local Area Personal Income. (2017). Available at: www.bea.gov.

83. U.S. Census Bureau. Population, Housing Units, Area, and Density. (2017). Available at: www.factfinder.census.gov.

84. U.S. Bureau of Labor Statistics. Local Area Unemployment Statistics. (2017). Available at: www.bls.gov.