Technical Appendix for "Enhancing equity while eliminating emissions in California's supply of transportation fuels"

Principal Investigators

Olivier Deschenes, Department of Economics, University of California, Santa Barbara
Ranjit Deshmukh, Environmental Studies Department, University of California, Santa Barbara
David Lea, Earth Science Department, University of California, Santa Barbara
Kyle Meng, Bren School and Department of Economics, University of California, Santa Barbara
Paige Weber, emLab, University of California, Santa Barbara and Department of Economics, University of North Carolina at Chapel Hill

Researchers

Tyler Cobian, emLab, University of California, Santa Barbara
Danae Hernandez Cortes, Department of Economics, University of California, Santa Barbara
Ruiwen Lee, emLab, University of California, Santa Barbara
Chris Malloy, Department of Economics, University of California, Santa Barbara
Tracey Mangin, emLab, University of California, Santa Barbara
Measrainsey Meng, emLab, University of California, Santa Barbara
Madeline Oliver, Bren School, University of California, Santa Barbara
Sandy Sum, Bren School, University of California, Santa Barbara
Vincent Thivierge, Bren School, University of California, Santa Barbara
Anagha Uppal, Department of Geography, University of California, Santa Barbara

Project Management

Michaela Clemence, emLab, University of California, Santa Barbara
Tia Kordell, emLab, University of California, Santa Barbara
Erin O'Reilly, emLab, University of California, Santa Barbara

December 2020

Contents

1 Data

1.1 Extraction segment

1.1.1 Production data

We obtain information on oil wells and their production history from California's Department of Conservation (DOC). We use several DOC datasets.

Crude oil production (in barrels) is recorded in the WellSTAR database (Department of Conservation, n.d.*c*). WellSTAR records well-level monthly production volumes for the 1977-2019 period. Wells in the United States are assigned a unique, permanent 12-digit identification number known as the API number, which WellSTAR tracks. The first 10 digits indicate the 10-digit well API number, which represents a surface-level hole. Wells typically start with a single wellbore, but additional wellbores can be drilled. The last two digits of the API number represent distinct wellbores or producing zones below the surface. In this study, we define a well by its unique 10-digit API number.

Each entry in the DOC's WellSTAR production dataset is associated with a three-digit field code. We are able to match this field code with a corresponding field name using the DOC's 2019 Wells dataset (Department of Conservation, n.d.*d*), which is a dataset of well-level identifiers and characteristic. We obtain spud date and 2019 well status (e.g., active, plugged, idle) from the DOC's All Wells dataset, which includes other well-level identifiers and characteristics (Department of Conservation, n.d.*a*). These datasets were obtained from California's Department of Conservation[1].

We exclude gas fields by removing entries corresponding to fields with the word " Gas" in their names (e.g., Cal Canal Gas). Even though crude oil is sometimes produced from gas fields, the proportion in California is small — less than 1% of oil production is from gas fields. More importantly, when oil is a by-product rather than the primary product, a well operator is more likely to respond to gas prices rather than oil prices. Therefore, it does not make sense to include gas wells in our model of well entry. We also exclude the field code "000" (i.e., "Any Field"), as this production is not associated with a specific field. Finally, we reassign historical production from the field "Old Wilmington" to "Wilmington", as field boundary and well location data suggest that the two fields are spatially equivalent — the spatial data containing field boundaries does not include a boundary for Old Wilmington. Exceptions to these processing procedures are noted where relevant.

For geolocation data, we turn to two datasets from the California DOC, Geologic Energy

[1]Although the DOC datasets used in this study were provided by the DOC, some of the data used is publicly available online: https://wellstar-public.conservation.ca.gov/General/Home/PublicLanding

Management Division. The first dataset contains the geographical boundaries for 516 oil and gas fields (Department of Conservation Division of Oil, Gas, and Geothermal Resources, n.d.). The second dataset contains the geographical coordinates for California oil and gas wells (California Department of Conservation, Geologic Energy Management Division, 2020). This dataset includes coordinates for all but 20 California wells that produced oil in the historic period.

1.1.2 Cost data

We construct variables for new well costs and oil field depletion from data obtained from Rystad Energy. Rystad Energy provides data for each oil-producing "asset" in California. Each asset is a group of wells operated by usually one but sometimes several companies within a geographic area. The dataset begins in 1900 and is available through the present. The specific year in which data for each asset starts depends on when the asset was first explored or developed.

We use Rystad's estimates for the capital expense (CapEx) and operating expense (OpEx) for each asset that produces oil or condensate in California (Rystad Energy, 2020*c*). Operational costs (OpEx) include production expenditures, transportation expenditures, staff and administrative costs, taxes, and abandonment costs. Capital expenditures (CapEx) include all development costs related to facilities and drilling wells. Conversations with Rystad staff reveal that these are estimates of realized costs paid by operators, which reflect the market price of producing from the respective assets. Both CapEx and OpEx are in dollars per barrel of crude produced ($/bbl) and thus represent the average CapEx and OpEx cost for a given asset.

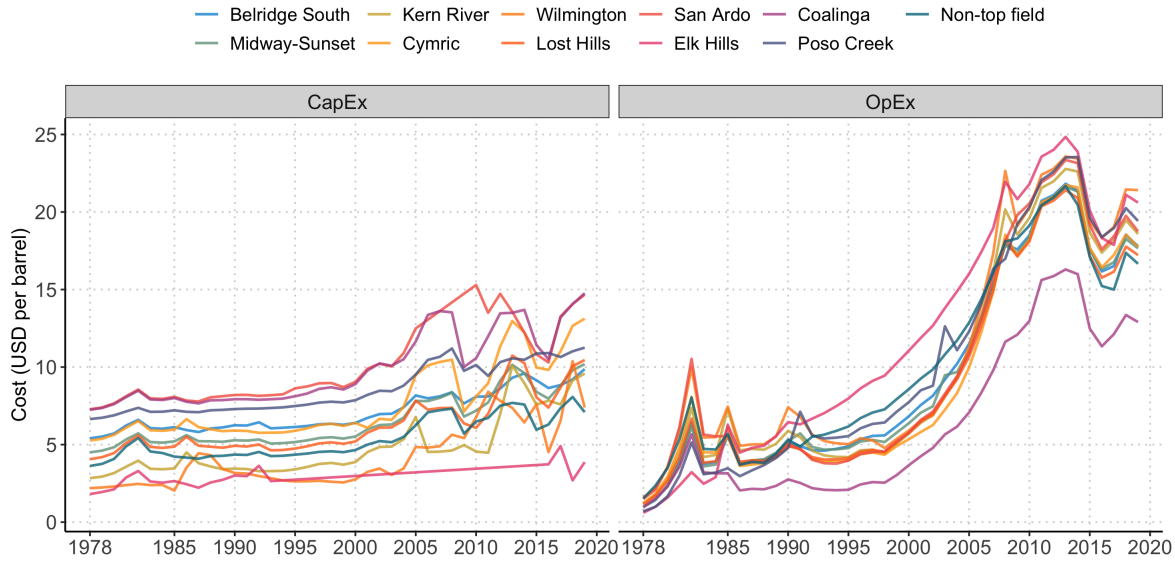
CapEx and OpEx values from the Rystad data are missing in some years for some assets within their sample period, requiring us to impute missing cost values. Furthermore, some CapEx and OpEx values are reported as zero. Since costs are unlikely to be zero and because the cost variables are Rystad's estimates for realized costs, we interpret these as years in which operators did not have to pay the cost rather than the price of producing being zero. Hence, cost observations with zero values are treated as missing and imputed as well. Because distributions of both observed CapEx and OpEx are skewed-right, we treat large outliers as missing (those above the 99^{th} percentile for CapEx and those above the 99.5^{th} percentile for OpEx) and also impute them. To impute the cost variables' missing values during the historic period, we perform linear interpolation using the non-missing values for each asset's CapEx and each asset's OpEx separately.

Assets do not correspond directly to a DOC-defined oil fields and an asset may not fall

within the spatial boundaries of a single DOC-defined oil field. To address this, we assign asset-level cost variables to fields using a two-step matching process. Rystad provides some information regarding which wells belong to a given asset (Rystad Energy, 2020*a*). For such wells, we clean the API number to reflect the ten-digit API number also used in the DOC data. In the first step of the matching process, we join Rystad's well information with the DOC's "All Wells" dataset, matching on ten-digit API (note that gas fields are not excluded in this step). We assume that an asset is matched to a field if it contains at least one matching well and we allow multiple assets to match to a single field. For fields that are not matched to any asset in the first step, we use the DOC field boundaries to match based on spatial proximity to a field that was successfully matched with at least one asset in step one. For each unmatched field, we find its nearest neighboring field(s) based on shortest distance between polygons that already has assets matched from the first step. The unmatched field is then assigned the same assets as its nearest neighbor(s). Because a field could be matched to multiple assets in either step, we calculate each field's CapEx as the mean of the matched assets' CapEx, and similarly, the field's OpEx is set to the mean of the matched assets' OpEx.

Figure 1 shows the historical series for CapEx (left panel) and OpEx (right panel) for each top field category after the imputation process. The costs of the non-topfields are averaged across fields in each year. The figures show that both CapEx and OpEx trend upwards over the years, with OpEx rising more steeply than CapEx.

Figure 1: Historical CapEx and OpEx (1978-2019)

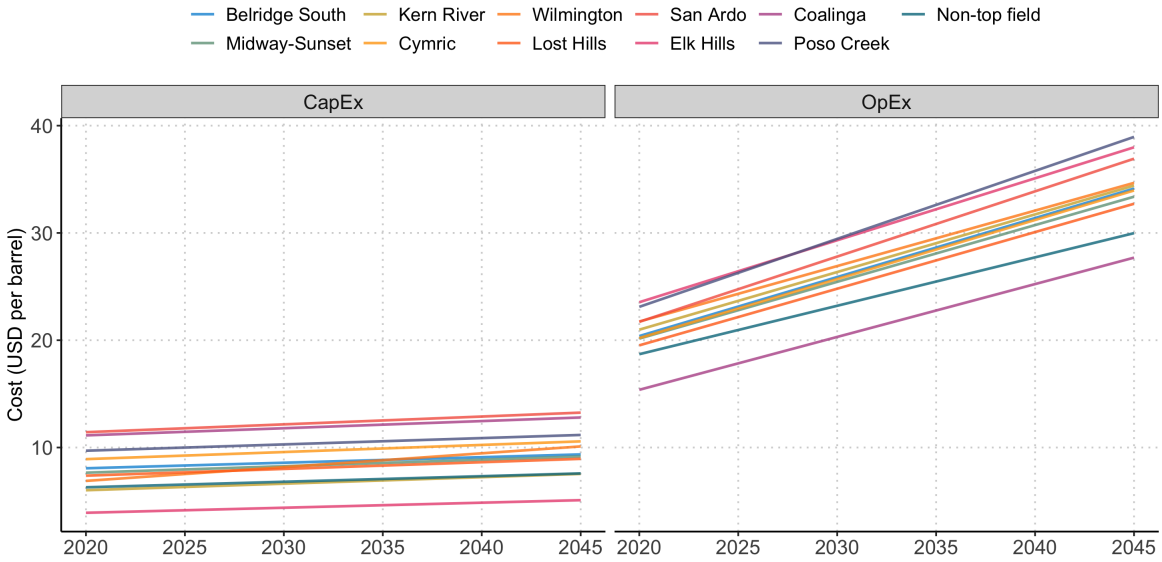


NOTES: Final time series of capital expenditure per barrel (left panel) and operating expense per barrel (right panel) in the historic period after missing values are imputed using linear interpolation. The eleven lines represent costs for each of the ten top fields and the average across the non-top fields.

To forecast well entry, values for the cost variables are needed from 2020-2045. We extrapolate each asset's CapEx and OpEx historical series using linear regression in time to forecast its future values. However, some historical cost series have too few observations for reliable extrapolation on their own. We thus pool any CapEx series that has less than twenty years of non-missing historic values and use a linear regression model with asset fixed effects to extrapolate the future CapEx for these assets. The same strategy is used for OpEx series with less than twenty historic observations.

Figure 2 shows the CapEx (left panel) and OpEx (right panel) series projected for the forecast period. The costs of the non-topfields are averaged across fields in each year. The figures show that both CapEx and OpEx trend upwards over the years, with OpEx rising more steeply than CapEx.

Figure 2: Projected CapEx and OpEx (2020-2045)



NOTES: Projected time series of capital expenditure per barrel (left panel) and operating expense per barrel (right panel) in the forecast period based on linear extrapolation from the historic series.The eleven lines represent costs for each of the ten top fields and the average across the non-top fields.

1.1.3 Historic well entry

We use the DOC production dataset to determine historic well entry in each field. Ideally, we would use each well's spud date. However, the spud date data that we were able to obtain is unreliable and incomplete. Of the 131,380 wells that according to the WellSTAR monthly production data produced oil during the historic period 1977-2019, only 27 percent were associated with one spud date, and 258 productive wells were associated with multiple spud dates. We therefore rely on the DOC's production data for determining the age and production start date of wells. We assume that the first recorded non-zero production date after 1977 for each well represents its entry date. We exclude production data from 1977 because it is in the first year of recorded production — thus, we are unable to distinguish between wells that began production in 1977 from those that began production before 1977. We assume that a well's production start date is the first non-zero production date after 1977 reported in WellSTAR, although in reality the true start date could be before 1978. We track annual entry by field. Some wells have recorded production in multiple fields. These wells are assigned to the first field in which they produced oil. Using this information, we are able to aggregate at the field-year level to determine historic well entry in each field annually during 1978-2019.

1.1.4 Field depletion data

Field-level depletion data is not available so we construct it using other data sources. Ideally, the field-level depletion variable would be constructed by dividing each field's cumulative production in each year by its constant technically recoverable resource. However, we have neither the actual numerator nor denominator values. We thus construct the variable first for the historic period using the best proxies we can find.

In the historic period, the depletion variable is constructed from Rystad data. Rystad provides each asset's historical oil and condensate production from 1900 onwards (Rystad Energy, 2020*e*). We use this data to calculate each asset's cumulative production in each year. Rystad also estimates the maximum economically recoverable resource given a high constant future Brent price of $120, which we use as a proxy for an asset's technically recoverable resource (Rystad Energy, 2020*d*). We calculate the historical asset depletion as cumulative production divided by recoverable resource:

$$d_{b,t\leq 2019} = \frac{\sum_{s=1900}^{t} q_{b,s}}{ERR_b}$$

where b denotes assets and t denotes years in the historical period. d is depletion, q is annual production, and ERR is an asset's economically recoverable resource. These historical asset-level depletion values are assigned to fields using the same matching algorithm described at the end of Section 1.1.2. The idea is that even though a field's production volumes and quantity of recoverable resource cannot be directly obtained from those of its matched assets in a reliable way, a field's depletion is best approximated from the depletion of its matched assets.

For the forecast period, however, a different strategy is required. While Rystad estimates each asset's production in the future, the estimates are based on a fixed set of Rystad's own assumptions, including future oil prices and future costs, which we cannot vary to create the scenarios we want to explore (details on the macroeconomic conditions and policy levers are respectively in Section 8 and Section 7. We thus take a different approach. First, we back out the field-level technically recoverable resource (TRR) from each field's estimated historical depletion described above and each field's historical production since 1977 from DOC data. This then allows us to calculate each field's depletion by dividing its cumulative production modeled in the forecast period by its constant TRR. We back out each field's TRR by first defining field depletion as a field's cumulative production divided by its recoverable resource:

$$d_{f,t\leq 2019} = \frac{\sum_{s=1900}^{t} q_{f,s}}{TRR_f}$$

where f denotes fields and t denotes years in the historical period. As before, d is depletion and q is oil production. TRR is a field's technically recoverable resource. Since field-level oil production data from DOC is available only from 1977 onwards, we split the years of production into two parts, before 1977 and 1977-2019:

$$d_{f,t\leq 2019} = \frac{\sum_{s=1900}^{1976} q_{f,s} + \sum_{s=1977}^{t} q_{f,s}}{TRR_f}$$

The first part is a field's cumulative production up to 1976 so it is a constant, which we call A below. The second part is a field's cumulative production from 1977 onwards and is calculated directly from DOC production data.

$$d_{f,t\leq 2019} = A_f + \frac{\sum_{s=1977}^{t} q_{f,s}}{TRR_f}$$

Writing cumulative production $\sum_{s=1977}^{t} q_{f,s}$ as Q_t, we arrive at the following equation to estimate TRR:

$$d_{f,t\leq 2019} = A_f + \frac{1}{TRR} Q_{f,t}$$

Since d is actually constructed with Rystad's ERR as the denominator instead of TRR, and ERR is a function of TRR and oil price, we modify the equation as follows:

$$d_{f,t\leq 2019} = A_f + \frac{1}{TRR} \frac{Q_{f,t}}{Brent_t}$$

where $Brent$ is the Brent benchmark price. We estimate this equation as a linear model separately for each field. The inverse of the coefficient on $\frac{Q_{f,t}}{Brent_t}$ yields the estimate for TRR. Field depletion in each future year $t > 2019$ is then calculated as:

$$d_{f,t} = d_{f,t-1} + \frac{q_{f,t}}{\widehat{TRR_f}} \tag{1}$$

1.2 Refinery segment

1.2.1 Refinery identifiers and location

The EIA provides information on existing refinery names, location, and capacity (Energy Information Administration, 2020*c*). For potential future installments of renewable refineries or refinery conversions, we collect data from a variety of sources such as press releases, news articles, etc. The following planned refinery conversions and installments are incorporated into our refinery modeling:

- Phillips 66 Rodeo San Francisco refinery separate unit (2021) – 8,000 barrels per day of renewable capacity installed (Tuttle, 2020)
- Marathon Petroleum Golden Eagle Martinez refinery conversion (2022) – 161,500 barrels per day of crude capacity shut down and 48,000 barrels per day of renewable capacity installed (Tuttle, 2020)
- Global Clean Energy (2022) – 15,000 barrels per day of renewable capacity installed (ExxonMobil, 2020)
- Phillips 66 Rodeo San Francisco refinery conversion (2024) – 75,700 barrels per day of crude capacity shut down and 44,357 barrels per day of renewable capacity installed (Phillips 66, 2020)

Our refinery model includes the Phillips 66 Santa Maria refinery and Phillips 66 Rodeo refinery with a total of 120,000 bpd capacity. The model assumes that only the capacity attributed to the Rodeo facility (76,000 bpd) is converted for refining of renewable fuels and that the Santa Maria unit would remain (about 45,000 bpd capacity) producing conventional fuels. During the comment period the authors were made aware that the Santa Maria unit may also convert to produce renewable fuels. As we have not captured this, if the Santa Maria unit does convert to renewable fuel, our model overestimates the remaining conventional crude capacity by the capacity of this Santa Maria unit.

Our refinery model does not include the conversion and expansion of Alt Air Paramount to produce renewable fuels, which the authors were made aware of during the comment period.

1.2.2 Refinery inputs and outputs

We utilize crude consumption and refined products (gasoline, diesel, and jet fuel) production data from the CEC Weekly Fuels Watch report (California Energy Commission, 2020). The dataset includes weekly reports of crude oil input and fuel (i.e., gasoline, diesel, and jet) production at the cluster (i.e., North and South) level. Renewable diesel consumption is obtained from a dataset provided by the CEC. We use data on refined product imports and exports from the CEC's Finished Products Movement dataset (California Energy Commission , CEC). Energy intensities (in units of MMBtu per bbl) for each fuel are taken from the EIA (Energy Information Administration, 2020*b*).

1.2.3 Fuel demand projections

Crude consumption at California's refineries depends on demand for refined fuels (gasoline, diesel and jet fuel) from within the state of California, demand for jet fuel for inter-state aviation transport, and exports of refined fuels out of California. Further, refineries can also produce renewable fuels (renewable gasoline, diesel and jet fuel) by either co-processing renewable feedstocks with crude oil or converting completely to process renewable feedstocks. Both the capacity utilization factor (operating level of a refinery compared to its rated processing capacity) and GHG emissions depend on the total production of refined fuels from a refinery.

Study 1 from the The Institute of Transportation Studies at UC Davis (ITS-Davis) projects in-state refined fuels demand under two scenarios: Business-As-Usual (BAU) and Low Carbon 1 (LC1). Out of all fuels that Study 1 projects for California's transportation sector, we use projections for the following fuels as inputs to the refining model: gasoline, diesel, drop-in gasoline (renewable gasoline), renewable diesel and sustainable aviation fuel (renewable jet fuel). ITS-Davis also projects jet fuel demand, but only for intrastate aviation transport. However, California refineries are responsible for producing jet fuel for both intrastate and interstate aviation transport. Thus, instead of Study 1 projections, we adopt the CEC's projections for total jet fuel demand (under the Mid Case scenario) until 2030 (Bahrenian et al., 2017), which accounts for both intrastate and interstate aviation transport and extrapolate it to 2045. We assume this same jet fuel demand projection for both the BAU and LC1 scenarios. Lastly, we account for military jet fuel demand, which is excluded from the CEC's demand projection, by adding the average annual military jet fuel demand from 2004 to 2012 uniformly to the CEC-based annual jet fuel demand projection.

Exports of refined products (gasoline, diesel, and jet fuel) are also included in our fuel demand projections. We add net exports of refined products (California Energy Commission , CEC) on top of the in-state refined fuel demand (the amounts of net exports assumed are dependent on the scenario, as explained in 7). In an earlier iteration of our model, we added total exports (instead of net exports) as part of fuel demand, resulting in greater total fuel demand. These original outputs were used in the modeling of the health and labor impacts. We have since corrected our refining model outputs to use net exports instead of total exports. To adjust the modeled health and labor impacts as well, we calculate linear adjustments of crude consumed and refined fuel produced (at the individual refinery level and at the county level for every year) using net exports compared to using total exports.

1.2.4 Crack spread for refined products

The prices received by refiners for refined products are projected for the forecast period by adding their historic spreads over Brent in real terms to the Brent crude oil price. For each refined product, the spread in a given year is the difference between the product's real annual wholesale price in California and the real Brent price in the same year. We collect the nominal annual wholesale prices over the ten years between 2005-2014[2] and the nominal annual Brent spot price from the EIA (Energy Information Administration, 2020*e*,*d*). We then inflate these price data to 2019 dollars using the consumer price index[3]. These prices in 2019$ are then inflated to 2020$ based on our assumption of inflation of 2% annually during the forecast period. We calculate the average real spread in 2020$ between 2005-2014 separately for each refined product:

$$Spread_k^R = \frac{1}{10} \sum_{t=2005}^{2014} (p_{k,t}^R - Brent_t^R)$$

where k denotes the three refined products: gasoline, diesel and jet fuel. $Spread^R$ is the average real annual spread, p^R is the real wholesale price of each refined product, and $Brent^R$ is the real annual Brent benchmark price; all variables are in 2020 dollars.

Table 1: Average real crack spreads between 2005-2014

Product	Crack spread (2020 USD)
Motor gasoline	23
Jet fuel	20
Diesel	23

NOTES: Average crack spread — the difference between the price received by refiners for finished products and the price refiners pay for crude oil (approximated with the Brent benchmark price) — for each of the three major types of refined petroleum products in real terms (2020 dollars).

1.3 Historic crude oil prices

We build our model of oil well entry using the annual Brent benchmark crude oil price over the historic period. The historic annual Brent data series from 1978-2019 is obtained from Rystad Energy (Rystad Energy, 2020*b*).

[2]We limited the time span to years before 2015 because a refinery-explosion related refined products shortage caused the spread between California's refined products and Brent crude oil price to rise. Although the refining capacity was back online the following year eliminating the shortages, the spread did not decrease for unknown reasons (Borenstein, 2020).

[3]Based on all items in U.S. city average, all urban consumers, and not seasonally adjusted, with a 1982-1984 base period obtained from the US Bureau of Labor Statistics (Bureau of Labor Statistics, 2020)

1.4 Health data

We perform the health analysis by calculating the emissions from total barrels refined and extracted at each site by multiplying total production by emission factors. We linked these emissions to facilities characteristics to obtain exposure at the census tract level. We also used census tract-level population characteristics to calculate health exposure to refining and extraction sectors. The following sections describe the underlying data for the health analysis.

1.4.1 Facilities characteristics

We obtained information on facilities characteristics directly from CARB. This information provides data on the stack height, diameter, velocity, and temperature of the refineries in the study. In case that there is more than one stack, CARB provided us with the average value of these variables. For the refinery facilities that were missing stack heights, we used the average characteristics of the refineries in the refinery cluster. We linked this information to the refineries location in order to calculate pollution exposure at the census tract.

1.4.2 Population data

Population data at the census tract was obtained from the 2010 Census. We calculated population projections for 2020-2045 by using age-group specific population growth rates at the county level from the California Department of Finance population projections.

1.4.3 Health incidence data

We obtain baseline incidence data for mortality and morbidity endpoints from BenMAP. For mortality incidence between 2019 to 2045, BenMAP uses both the U.S. Centers for Disease Control (CDC) WONDER database and the Census Bureau national mortality rate tables. Morbidity incidence rates are from the Healthcare Cost and Utilization Project (HCUP). Baseline morbidity rates from HCUP are for 2014. Specifically, we consider four morbidity outcomes: cardiovascular hospitalizations (International Classification of Disease ninth revision (ICD-9): 390-429), respiratory hospitalizations (ICD-9: 460-519), non-fatal heart attacks (ICD-9: 410), and emergency room visits for asthma. The first three morbidity outcomes represent hospital admissions from the emergency room, and the last outcome daily emergency room visits. Since asthma is a subset of respiratory hospitalizations, double counting could occur if an emergency room visit for asthma-related conditions resulted in hospitalization.

1.4.4 Income growth rates

Monetized values of the health benefits in future periods are adjusted according to income growth rates using national output projections from the Congressional Budget Office and the income elasticities used in BenMAP.

1.4.5 Disadvantaged population data

We obtained information on the population living in disadvantaged communities from the CalEnviroScreen 3.0 (CES) developed by the Office of Environmental Health Hazard Assessment (OEHHA). CES scores each census tract in California according to environmental variables and population characteristics to identify communities that are both disproportionately burdened with high pollution and are vulnerable to pollution exposure. The CES score is obtained by quantifying the pollution exposure, health vulnerability, and socioeconomic characteristics and creating a score for each category. A census tract is designated disadvantaged if one of the two following conditions hold: (1) its score exceeds the 75th percentile of statewide scores, or (2) areas that are among the top 5% pollution burden without reliable population characteristics.

1.5 Labor data

We perform the labor analysis using proprietary data from IMPLAN. To construct its underlying data, IMPLAN draws on over 90 sources of information including the Quarterly Census of Employment and Wages (QCEW) from the Bureau of Labor Statistics, County Business Patterns from the Census Bureau, and the Regional Economic Accounts (REA) from the Bureau of Economic Analysis.

1.5.1 Inter-Industry Linkages

IMPLAN bases it's input-output model on the Benchmark Input/Output Tables released by the Bureau of Economic Analysis every 5 years.

1.5.2 Employment

IMPLAN constructs measures of employment primarily using the QCEW and supplementing with the CBP whenever employment values in the QCEW are undisclosed. Furthermore, since the QCEW only covers firms which participate in federal unemployment insurance programs, IMPLAN uses the REA to estimate the amount of proprietorship employment

in each industry at the county level. For more information on IMPLAN's construction of employment data see IMPLAN Group (2020).

1.5.3 Worker Compensation

IMPLAN uses the QCEW to obtain measures of employee compensation for each county and 6 digit NAICS industry. Wherever compensation data is not disclosed in the QCEW, IMPLAN uses the REA to provide supplementary information. For more information on IMPLAN's construction of compensation data see IMPLAN Group (2020).

1.5.4 Data for Multi-Regional Input Output Analysis

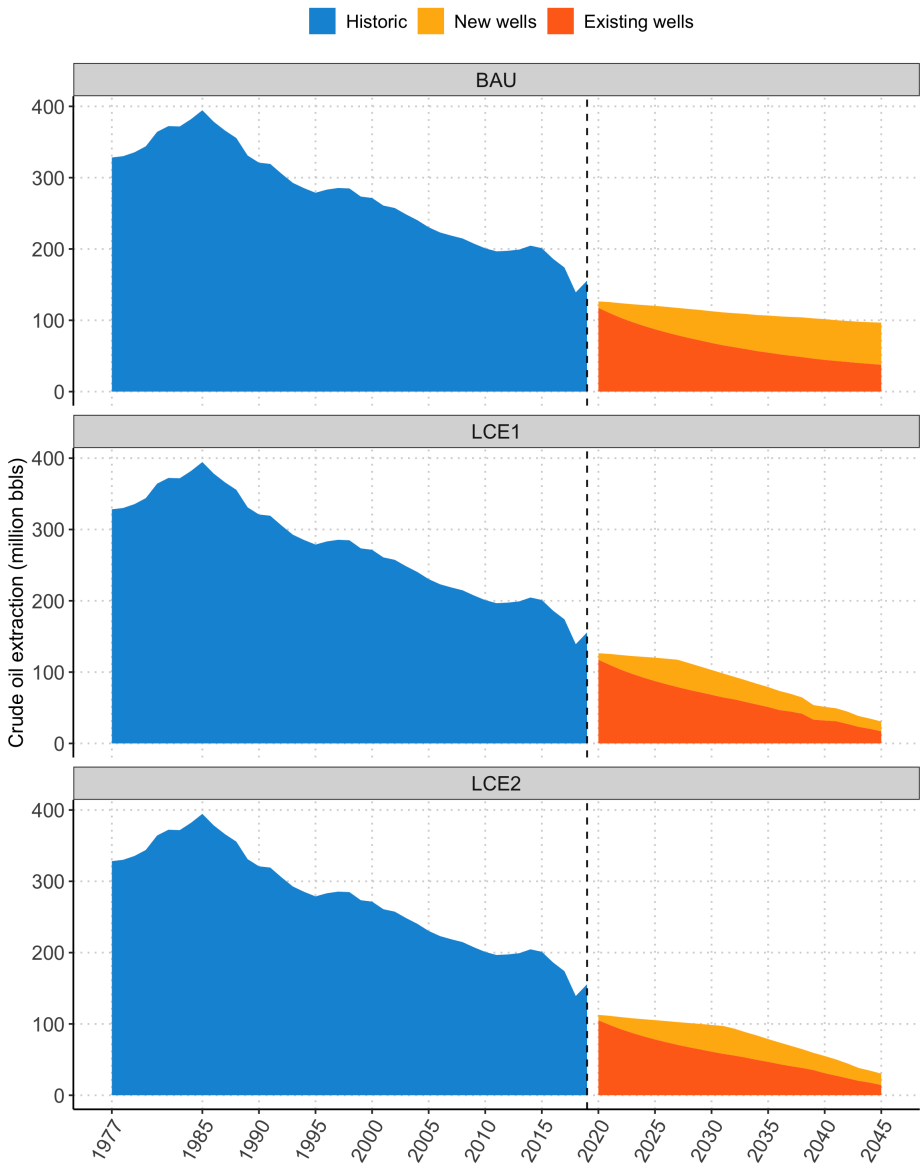
IMPLAN uses information from multiple data sources to capture inter-regional supply chain linkages and spending patterns. Inter-regional linkages are modeled using a gravity-model of trade which is calibrated using information on commodity trade and transportation costs from the Commodity Flow Survey (Census), Freight Analysis Framework (Federal Highway Transportation Administration), and an integrated, intermodal transportation network modeling system (Oak Ridge National Laboratory). For more information on IMPLAN's gravity model of trade and related data sources, refer to Squibb and Thorvaldson (2020).

2 Model of crude oil extraction

California's future crude oil supply comes from a combination of existing in-state wells, newly drilled in-state wells, and imports from out-of-state. To simplify the analysis, we do not model out-of-state oil supply curves.[4] Instead, we assume that California refineries can import crude oil from out-of-state exporters whose supply is perfectly elastic. Our in-state model of crude oil production includes three components. Section 2.1 details a statistical model that predicts entry of new wells. Section 2.2 discusses a deterministic model that predicts production from wells after they enter. Finally, we apply a rule dictating the exit of wells in Section 2.3. The resulting production forecast by our model under the three scenarios examined in the main report (see Table IV.1) is shown in Figure 3.

[4]Doing so would require modeling the oil production decisions of every oil producing nation in the world out to 2045. This is beyond the scope of our study.

Figure 3: Oil production from existing and new wells



NOTES: Historic production (blue area) and future production from existing wells (orange) and from wells that enter in the forecast period (yellow) under the three scenarios examined in the main report, BAU (top panel), LCE1 (middle) and LCE2 (bottom).

2.1 Entry of new oil wells

Our model for oil well entry is motivated by the principle that a well is drilled in an oil field when it is profitable to do so. Unfortunately, there is a major empirical hurdle in modeling new crude oil drilling decisions: we do not observe the profitability (investment and operations costs) of all potential oil well sites. Instead, our approach is to use available

cost data together with historical global oil prices to empirically estimate the relationship between historical well entry and proxy variables that capture determinants of oil production profitability during the period between 1978-2019. Extraction cost data come from the subscription-based data provider, Rystad Energy, as well as from the DOC. We then use these historical estimates to project the entry of oil wells into production over the forecast period between 2020-2045.

2.1.1 Model estimation

The crude oil production decision across oil fields in California is a function of expected revenue and costs for all existing and potential oil production sites across the state. We do not directly observe these variables, particularly for potential crude oil production sites. Instead, we use observable costs, global crude oil prices, and historical entry data to statistically estimate how costs and prices have impacted historical entry over the historic period 1978-2019. Insofar as observed proxy cost variables and current crude oil prices are correlated with determinants of new oil wells, our statistical model provides an empirical relationship between the variables that we do observe with production from new wells.

Rystad provides measures of operational costs and capital expenditures at the asset-level. Most assets in Rystad's data are groups of wells operated by one company in a contiguous geographic area but some assets could contain wells operated by several companies or they could cover a large geographic area. Operational costs (OpEx) include production expenditures, transportation expenditures, staff and administrative costs, taxes, and abandonment costs. Capital expenditures (CapEx) include all development costs related to facilities and drilling wells. Exploration costs are not included, as we view exploration activity as a sunk cost at the point of the production entry decision. Rystad reports CapEx annually based on the year incurred, as in this setting capital expenditures may be incurred in multiple years, with the largest CapEx occurring in the early years of well development. Our main unit of analysis in this study is a field-year, so asset-level data provided by Rystad are converted to the field level (see Section 1.1.2 for conversion details). As a proxy for expected crude oil prices, we simply use current global crude oil prices.[5]

Recent research has identified that option value and uncertainty may also be important factors in the entry decision (Kellogg, 2014). Specifically, firms deciding whether to enter also consider the value of delaying entry when they expect future states of the world to have higher returns to investment, while uncertainty impacts their expectation of these future

[5]There is empirical research showing that oil consumers believe future oil prices are simply the same as the current oil price (Anderson et al., 2011) such that current prices may actually be expected prices and not simply a proxy. However, it is unclear if such expectations carry over to crude oil producers.

returns. While we do not deny the importance of option value and uncertainty in the entry decision, a fully dynamic model of investment that includes uncertainty and option value is beyond the scope of the study. More importantly, however, the goal of our model is prediction, and not necessarily to identify specific mechanisms behind the entry decision. The model fit diagnostics shown in Section 2.1.2 demonstrates that our parsimonious model is well suited for predicting field-level well entry, even without explicitly including option value and uncertainty.

We empirically estimate the relationship between the number of new wells in each field, the prevalent benchmark oil price, the cost of entry, and how depleted fields are. We use a Poisson model as the number of wells is a count variable with a highly skewed distribution. Entry in the various fields are likely to respond differently to the same oil price change because oil grade varies across fields. We thus allow the effect of oil price to vary for top producing fields. Specifically, we estimate the following field-by-year Poisson model:

$$NewWells_{f,t} = \alpha_g + \sum_{g=1}^{11} \beta_g G_g \times Brent_t + \gamma_c CapEx_{f,t} + \gamma_o OpEx_{f,t} + \delta Depl_{f,t} + \epsilon_{f,t} \quad (2)$$

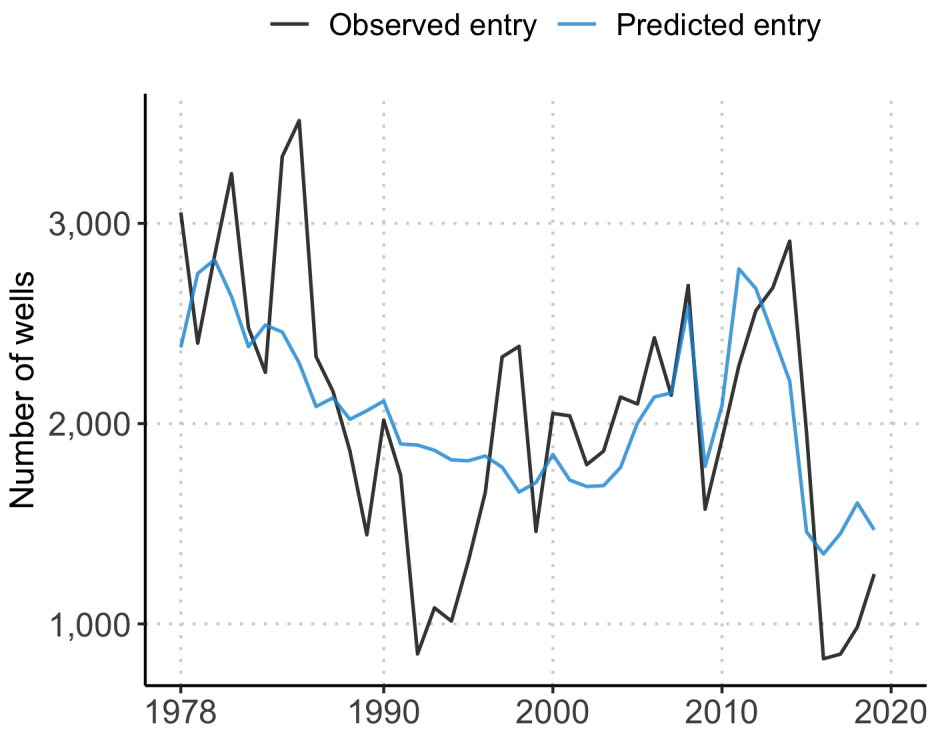
where f denotes each oil field, g denotes each top field category, t denotes year, and $NewWells$ is the number of new wells, constructed from DOC data (see Section 1.1.3 for details). G_g is a set of eleven indicator variables that represent each of the ten fields that produced the most oil in 2019 and a single category for the other 252 non-top fields.[6] α_g is a set of eleven top field-category fixed effects. $Brent$ is the Brent benchmark oil price in nominal USD per barrel. $CapEx$ and $OpEx$ are respectively the capital expenditure and operating expense of wells in each field, constructed from Rystad data (see Section 1.1.2 for details). $CapEx$ and $OpEx$ are both normalized to nominal USD per barrel. $Depl$ is a proxy for how depleted an oil field is (see Section 1.1.4 for details). It takes values in [0,1], where 0 indicates no oil has yet been produced from the given field and 1 indicates that the field is completely depleted. The top field-specific coefficients on $Brent$, the set of eleven β_g's, are separately estimated for each topfield category while the coefficients on $CapEx$ (γ_c), on $OpEx$ (γ_o), and on $Depl$ (δ) are jointly estimated. We estimate the model parameters using data from 1978-2019. $\epsilon_{f,t}$ is the error term.

[6]Insufficient data from non-top oil fields prevent us from estimating field-specific crude-oil price coefficients for those fields.

2.1.2 Model fit

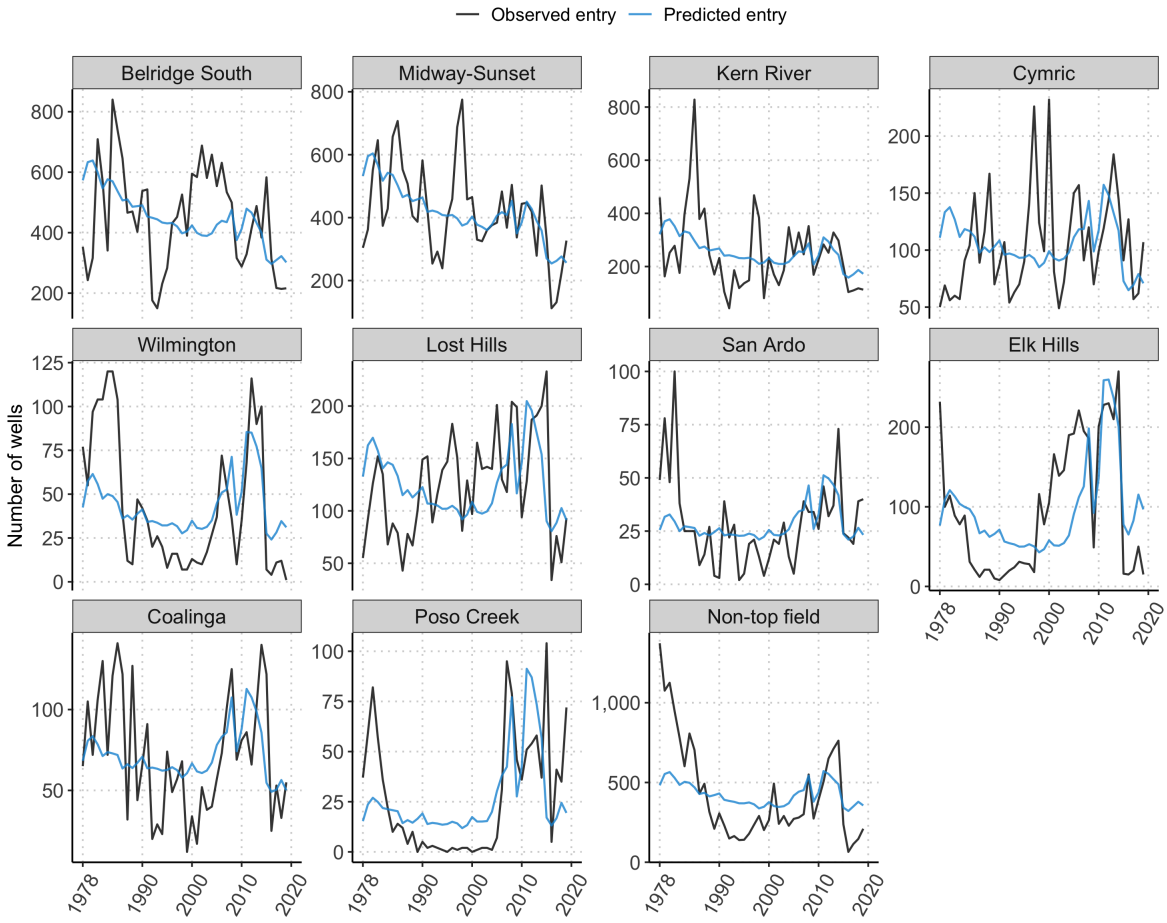
To verify that our model correctly predicts entry, we plot its prediction against observed entry over the historic estimation period. Figure 4 compares predicted and observed entry at the state level and Figure 5 compares the predicted and observed entry for each top field category. The plots indicate that the model predicts overall trends in entry quite well.

Figure 4: Predictions of new wells in California based on full historic dataset



NOTES: Total number of new wells in California — observed (black line) and predicted by our final model (blue line). Our final model uses the full historic dataset, i.e. data from 1978-2019, to predict well entry based on the parameters estimated in equation 2.

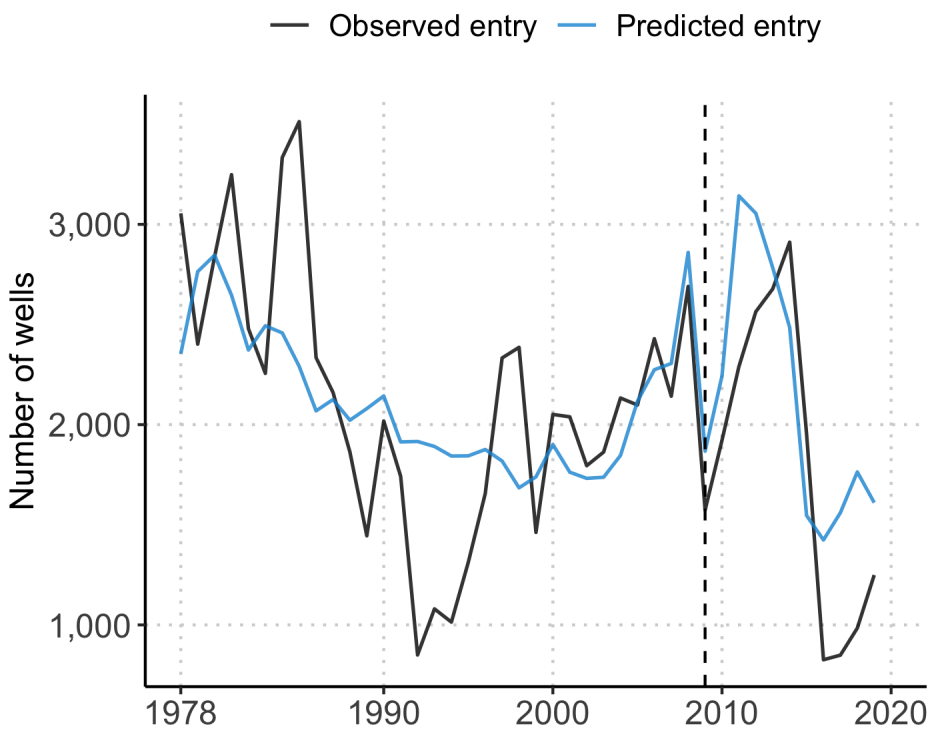
Figure 5: Predictions of new wells by top field category based on full historic dataset



NOTES: Number of new wells in each top field category — observed (black lines) and predicted by our final model (blue lines). Our final model uses the full historic dataset, i.e. data from 1978-2019, to predict well entry based on the parameters estimated in equation 2. The last panel (i.e., "Non-top field") represents the average well entry observed and predicted across all non-top fields.

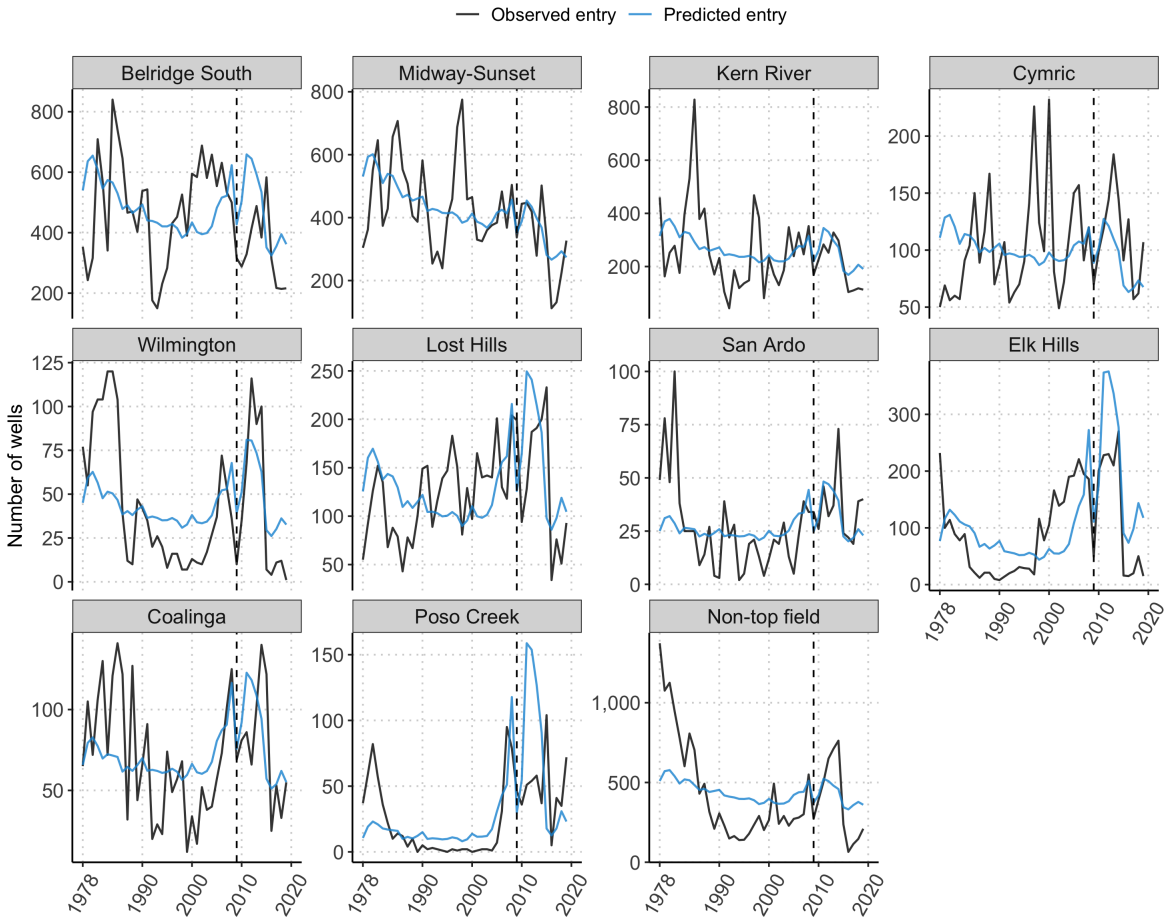
To check for out-of-sample fit, we re-estimate the model on data from 1978-2009, holding out data from 2010-2019 for out-of-sample validation. We re-plot predicted entry against observed entry for the entire historic period. Figure 6 and Figure 7 show that this modified model predicts entry trends well at the state and top field levels, respectively.

Figure 6: Predictions of new wells in California based on data from 1978-2009



NOTES: Total number of new wells in California — observed (black line) and predicted by a modified model (blue line). This modified model uses historic data from 1978-2009; data from 2010-2019 is held out for out-of-sample validation. Using this smaller dataset, the modified model predicts well entry based on the parameters estimated in equation 2.

Figure 7: Predictions of new wells by top field category based on data from 1978-2009



NOTES: Number of new wells in each top field category — observed (black line) and predicted by a modified model (blue line). This modified model uses historic data from 1978-2009; data from 2010-2019 is held out for out-of-sample validation. Using this smaller dataset, the modified model predicts well entry based on the parameters estimated in equation 2. The last panel (i.e., "Non-top field") represents the average well entry observed and predicted across all non-top fields.

2.1.3 Projecting future entry

To project entry during the 2020-2045 period, we use the statistical parameters from equation (2) and apply forecasted values for each regressor. The time series of projected oil price depends on the future global crude oil price scenario (see Section 8.2). Projected extraction costs are extrapolated from historic ones (see Section 1.1.2 for details). Depletion is calculated each year according to equation (1) (see Section 1.1.4 for details). Once depletion reaches 0.99 in a field, new entry is no longer allowed.

2.2 Production decline of oil wells

2.2.1 Decline parameterization

To model annual production from wells (both those that have entered in the historic period and those forecasted to enter in the future), we fit production decline curves similar to those used in the industry and academic literature to historical production data to estimate production decline parameters. We then apply the estimated parameters to the forecast years.

Wells that began production in the historic period are grouped by the field in which they are located and which five-year period they first began producing oil, which we denote as the well's vintage. Each vintage (except for the most recent vintage) is a five-year window. We use the following five-year periods as vintages: 1978-1982, 1983-1987, 1988-1992, 1993-1997, 1998-2002, 2003-2007, 2008-2012, and 2013-2019. We do not know the age of wells that have production data available in 1977, so we classify them as a "pre-1978" vintage. As done by the EIA (Energy Information Administration, 2020*f*), we consider production from an oil well as first experiencing hyperbolic decline followed by exponential decline as the well ages. Accordingly, we estimate decline parameters separately for each field-vintage by fitting a hyperbolic decline curve that intersects with an exponential decline curve to the wells' production data over time. For certain field-vintages, it is not possible for their production data to be fitted with a hyperbolic decline curve followed by an exponential decline curve. For example, the more recent vintages do not have enough data points. In these cases, we fit only an exponential decline curve.

We aggregate production data for each well to the annual level, where each year is 12 months from the well's start month. If a well began operating before 1977, we treat January 1977 as its start month. We then aggregate production further to the field-vintage-year level. The following are the steps for how we fit production decline curves:

1. For every year of production within every vintage, we calculate the *production rate*, which we consider to be the average production per day, by dividing annual production by 365 days.

2. We calculate the *decline rate* by calculating the percent change in production rate within a year from the previous year's decline rate:

$$Decline\ rate_t = \frac{Production\ rate_t - Production\ rate_{t-1}}{Production\ rate_t} \tag{3}$$

3. We determine the year number where production within a field-vintage peaks. We

then remove all years prior to this peak year.

4. We calculate a *moving decline rate* by taking a 5-year left-aligned rolling mean of the decline rate each year. We use this rolling mean as a way to estimate what the decline rate looks like within a year and for the subsequent 4 years, as there can be noise year-from-year when looking at just decline rates.

5. We estimate the year in each the productive curve changes from hyperbolic to exponential (t_{exp}) as the first year in which the moving decline rate hits below 10%. However, if such a year does not exist (as in, the moving decline rate within a vintage never reaches below 10%) or if the year fitting the criteria happens before the vintage's 30^{th} percentile year, we estimate that the field-vintage's production curve does not change from hyperbolic to exponential. The reason we use the 30^{th} percentile year is if the estimated year happens too early, there is not a sufficient number of data years to fit a hyperbolic curve.

6. We first fit an exponential curve for the entire span of the field-vintage's production curve (post-peak year).

7. We then test if a hyperbolic curve can be fit at all using the full post-peak year field-vintage data group. The equation for a hyperbolic production rate is shown in equation (4).

$$Hyperbolic\ production\ rate = \frac{q_i}{(1 + bD_h t)^{1/b}} \tag{4}$$

   where q_i is the peak production rate, D_h is the initial decline rate, t is the year, and b is the hyperbolic decline parameter. We set q_i as the peak production rate calculated in step 1. To avoid noise in the first few years' decline rates (such as a positive decline rate), we use the greatest decline rate out of the first five years (starting from the peak production year) as the D_h. We use a non-linear least squares (NLS) regression to solve for b (using $b = 1$ as the starting point).

8. If a field-vintage's production curve is able to be fitted as a hyperbolic decline, we then attempt to fit both a hyperbolic curve and an exponential curve for the given field-vintage. This time, we first start by fitting a hyperbolic curve up to the year before the estimated t_{exp} in step 5. From the t_{exp} year and all subsequent years, we fit an exponential decline to the production curve. The exponential production rate is defined in equation (5).

$$Exponential\ production\ rate = q_i e^{(-dt)} \tag{5}$$

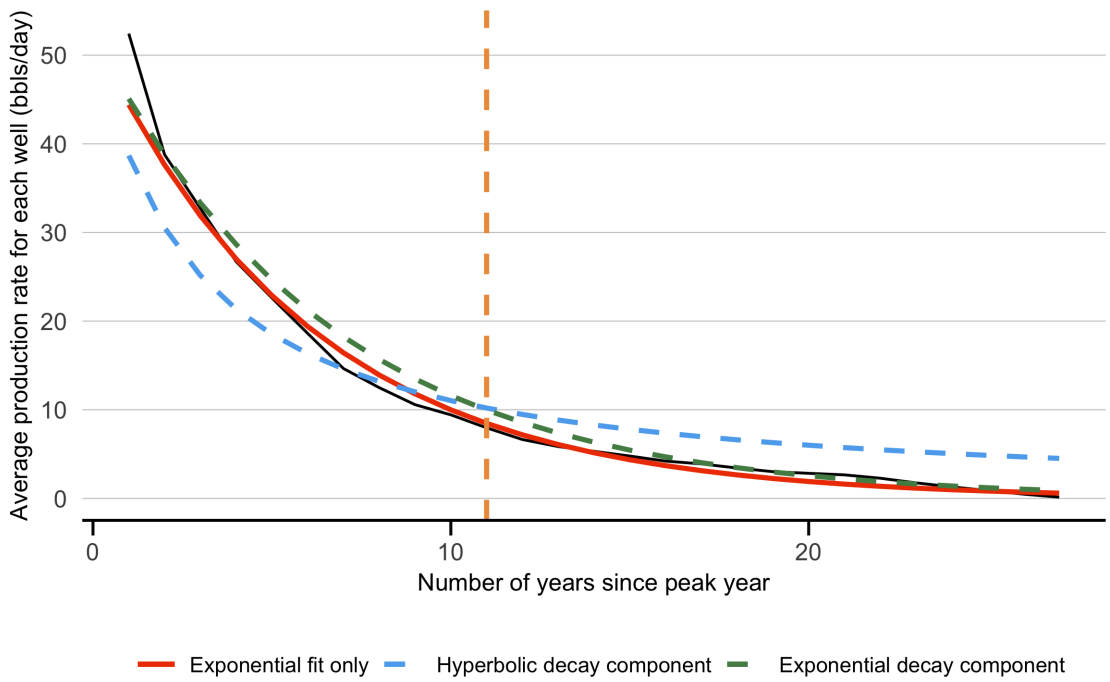
where d is the exponential decline parameter. We solve for d using NLS regression as well.

9. If it is not possible to fit both a hyperbolic and an exponential decline curve, we assume the field-vintage is defined by only a hyperbolic decline. If the field-vintage does have both a hyperbolic curve fit and an exponential curve fit, we use the year when the hyperbolic and the exponential curves intersect (i.e., when the difference between equation (4) and equation (5) is zero) as the intercept year. We solve for this intercept year by solving for the root of the difference of equation (4) and equation (5):

$$Intercept\ year = \frac{q_i}{(1 + bD_h t)^{1/b}} - q_i e^{(-dt)} == 0$$

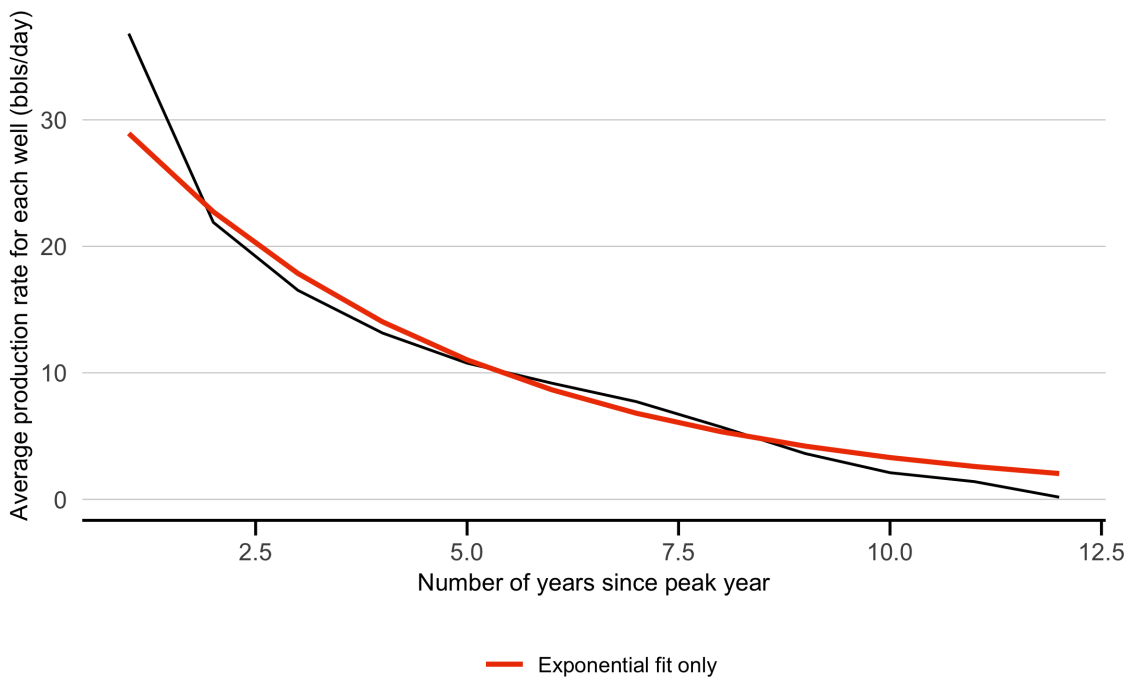
Figure 8 shows an example of wells whose production data is fitted with a hyperbolic curve followed by an exponential curve. Figure 9 shows another example of wells whose production data could not be fitted with both a hyperbolic curve and an exponential curve, so it was fitted with only an exponential curve.

Figure 8: Example of field-vintage fitted with hyperbolic and exponential curves



NOTES: Decline curve fitting to wells from the 1993-1997 vintage in the Belridge South oil field. The actual production rate (solid black line) for this group of wells could be fitted with a hyperbolic decay component (blue dashed line) and an exponential decay component (green dashed line). The orange dashed line shows the years since the peak production year in which the production profile switches from hyperbolic to exponential, i.e. where the blue dashed line intersects the green dashed line. The red solid line shows the curve fit if the production rate were to be fitted with exponential decay only.

Figure 9: Example of field-vintage fitted only with exponential curve



NOTES: Decline curve fitting to wells from the 2008-2012 vintage in the Elk Hills oil field. The actual production rate (solid black line) for this group of wells could not be fitted with both a hyperbolic decay component and an exponential decay component. Thus, it was fitted only with an exponential decay curve (solid red line).

2.2.2 Future production of existing wells

To model oil production from existing wells (wells that began production starting in or before 2019), we apply the decline parameters to each field-vintage. For each forecast year, we apply equations (4) and (5) to solve for a field-vintage's future production. The t value used is the difference between the forecast year and the field-vintage's start year. The q_i is the peak production rate (bbls/day) multiplied by the number of wells in the field-vintage.

2.2.3 Future production of new wells

Because the decline parameters b and d change over time, we extrapolate the historical parameters using linear regression to obtain values for the forecast period (2020-2045). The extrapolation is carried out separately for each of the top ten producing fields. For the remainder of the fields, we extrapolate using the median parameter values (across fields) in each vintage. The production-year in which the hyperbolic and exponential years intersect is similarly extrapolated from historical intersect years using linear regression.

2.3 Exit of oil wells

When average well production in time t falls below a respective field-specific threshold (in number of barrels), all of the wells within a field-vintage will exit and produce zero barrels of oil for the remaining years in the forecast period through 2045. Each field-specific exit threshold value is the average production of the field's plugged wells (i.e., those with the well status "Plugged" or "PluggedOnly") in each well's final production year. We define the final production year as the consecutive twelve months ending with a well's final month of non-zero production. Future production for all wells in a given field-vintage in time t is the same and is determined by the field-vintage decline curves detailed in Section 2.2). If the production value per well is lower than the field-specific well production threshold, the entire field-vintage does stops producing oil in time t and the subsequent years.

A limitation of our exit approach is that it is solely based on a production quantity. In reality, the exit of wells may be dependent on other factors such as oil price and well age.

3 Model of refinery production

The model of refinery production estimates crude consumed by refineries given an output mix of refined products and determines the operations, conversions (to refine renewable fuels), and shut downs of refineries. For refined products, we only consider crude-based gasoline, diesel, and jet fuel, and renewable gasoline, diesel, and jet fuel. Because the scenarios for refinery production influence the computational steps within the model, we present the scenarios first before presenting the model.

3.1 Crude consumption

Historical data on crude oil consumption and refined fuels production from California refineries is available at the level of two clusters – North and South. To estimate the relationship between crude consumption and refined fuels production at the refinery cluster levels, we utilize Equation 6:

$$V_{crude}EI_{crude} = V_{gasoline}EI_{gasoline} + V_{diesel}EI_{diesel} + V_{jet}EI_{jet} + \alpha V_{crude} \tag{6}$$

Where V_{crude} is the crude consumed in units of barrels (bbl). $V_{gasoline}$, V_{diesel}, and V_{jet} are the volumes of gasoline, diesel, and jet fuel demand in units of barrels, respectively. EI_{crude}, $EI_{gasoline}$, EI_{diesel}, and EI_{jet} are the energy intensities of crude, gasoline, diesel, and jet fuel in units of MMBtu per barrel, respectively. The α variable is a catch-all coefficient that

captures both additional refined products and efficiency losses at the refinery. Using annual historical data on crude consumption and fuel production from the CEC Weekly Fuels Watch Report (California Energy Commission, 2020), we solve for α within each region:

$$\alpha = \frac{V_{crude}EI_{crude} - V_{gasoline}EI_{gasoline} - V_{diesel}EI_{diesel} - V_{jet}EI_{jet}}{V_{crude}} \tag{7}$$

Using the historically-established relationship between refined fuel production and crude consumption, we then solve for the crude required to produce projected gasoline, diesel, and jet fuel demand using Equation 8.

$$V_{crude} = \frac{V_{gasoline}EI_{gasoline} + V_{diesel}EI_{diesel} + V_{jet}EI_{jet}}{EI_{crude} - \alpha} \tag{8}$$

3.2 Refinery operations

Crude oil consumption, estimated from refined products demand determines the operation levels for refineries. Depending on the operation levels, refineries either undergo conversions to produce renewable fuels, co-produce renewable fuels, or shut down if operation levels fall below profitable levels. We assume that renewable gasoline, diesel, and jet fuel have the same energy intensities as their crude-based counterparts. We allow only those refineries that have announced their conversions to completely convert to producing renewable fuels. For other refineries, we assume that they are able to process renewable feedstocks and co-produce renewable fuels alongside crude-based products. To simplify this model, we assume that refineries can produce gasoline, diesel, and jet fuel in any proportion and ignore the technical constraints on refinery product slates.

The following order of operations is applied to determine the operation levels, shut downs, installments, and conversion of refineries. For each year t within each scenario:

1. Estimate the total crude refinery capacity and renewable fuels refinery capacity (including announced conversions). If a conversion is to take place in that year, retire the crude refining capacity for that refinery and add announced renewable fuels processing capacity to the total renewable fuel processing capacity.

2. Convert renewable gasoline/jet/diesel (reGJD) demand into barrels gasoline equivalent (bge) using energy intensities of those fuels.

3. Split crude-based gasoline, jet, and diesel (GJD) demand by cluster (North and South) using historical (2015-2019) proportions of how much of each fuel is produced in each

cluster, as shown in Equation 9.

$$V_{k,c,t} = V_{k,t} \times p_{k,c} \tag{9}$$

Where $V_{k,t}$ is the demand for refined product k in year t and $p_{k,c}$ is the historic (2015-2019) average proportion of fuel k produced in cluster c.

4. Compare reGJD demand to converted renewable refineries capacity (announced conversions). If reGJD demand is less than 90% (assumed capacity utilization factor) of renewable refineries capacity, then no action (all reGJD demand will be produced by renewable refineries). If reGJD demand is greater than renewable refinery capacity:

   (a) If the Kern Oil refinery exists, assign up to 376,417,919 gallons of gasoline equivalent (gge) of renewable diesel to the refinery. The Kern Oil refinery co-produces renewable diesel along with crude-based refined products, and this amount is equal to its average historical production of renewable diesel in 2015-2019 California Air Resources Board (n.d.*c*).

   (b) Calculate residual reGJD demand in year t ($V_{k(ren),res,t}$) using Equation 10.

$$V_{k(ren),res,t} = V_{k(ren),t} - V_{Kern\ oil\ REdiesel} - B_{ren,t} \tag{10}$$

   Where $V_{k(ren),t}$ is the demand for renewable fuel $k(ren)$ in year t, $V_{Kern\ oil\ REdiesel}$ is Kern Oil refinery's renewable diesel capacity, and $B_{ren,t}$ is total renewable refinery capacity in year t. The residual reGJD demand $V_{k(ren),res,t}$ represents the amount of renewable fuels demand that cannot be met by existing renewable refinery capacity.

   (c) Split residual reGJD between North and South clusters using Equation 9 to obtain $V_{k(ren),res,c,t}$..We assume the residual renewable gasoline, renewable diesel, and renewable jet demand are split in the same proportion between the two clusters as crude-based gasoline, diesel, and jet fuel, respectively.

   (d) Add residual reGJD demand to crude-based GJD demand for North and South clusters to estimate total GJD demand for non-converted refineries in year t ($V_{k,c,t,tot}$), as shown in Equation 11.

$$V_{k,c,t,tot} = V_{k,c,t} + V_{k(ren),res,c,t} \tag{11}$$

   Where $V_{k,c,t}$ is demand for crude-based fuel k in cluster c and year t.

5. We define "equivalent crude demand" as a proxy for both crude oil and renewable feedstock consumed by refineries that either produce only crude-based products or co-produce both crude-based and refined products. We use Equation 8 to calculate the equivalent crude demand for crude refineries within each cluster, depending on the exports scenario:

    - For the "historic production" scenario: Set the crude demand within each cluster to meet the historic average capacity utilization factor (CUF) within each cluster. The difference between the total crude demand and the equivalent crude demand based on projected GJD demand is exported.
    - For the "historic exports" scenario: Add the historical average amount of gasoline, diesel, and jet fuel net exports to the total GJD demand. Calculate the equivalent crude demand.
    - For the "low exports" scenario: Add the amount of gasoline, diesel, and jet fuel that is assumed to be exported within that year to total GJD demand. Calculate the equivalent crude demand.

6. We assume that refineries are retired if their Capacity Utilization Factor (CUF) falls below 60%. Calculate the CUF within each cluster.

    - If CUF $>$ 1: Don't retire the crude capacity for refineries that have announced retirement/conversions. Add back the crude refinery capacity and subtract the renewable fuel processing capacity for that refinery.
    - If CUF $<$ 0.6: Retire the smallest refinery (We assume that the smallest refinery has the greatest marginal costs and hence, the first in line to retire). Recalculate CUF. If new CUF is greater than 1, don't retire that refinery. If CUF is still less than 0.6, repeat this step.
    - If CUF $>=$ 0.6 and CUF $<=$ 1: No additional steps needed.

3.3 Crude imports

After determining refinery operations and the amount of total demand (crude-based and renewable fuels) that need to be processed at crude refineries, we use Equation 8 to recalculate the amount of crude needed to meet crude-based refined product demand (gasoline, gasoline, and jet) only. We sum the crude amounts for the North and South clusters together and call this "actual crude oil demand" in year t, $V_{crude,actual,t}$. We can calculate the amount of crude

oil that must be imported into California by subtracting $V_{rude,actual,t}$ by the total amount of oil extracted in California oil fields in year t, as shown in Equation

$$q_{imported,t} = V_{crude,actual,t} - q_t \tag{12}$$

Where $q_{imported,t}$ is the amount of crude imported in year t and q_t is the total amount of oil produced in-state in year t.

4 Quantifying GHG emissions

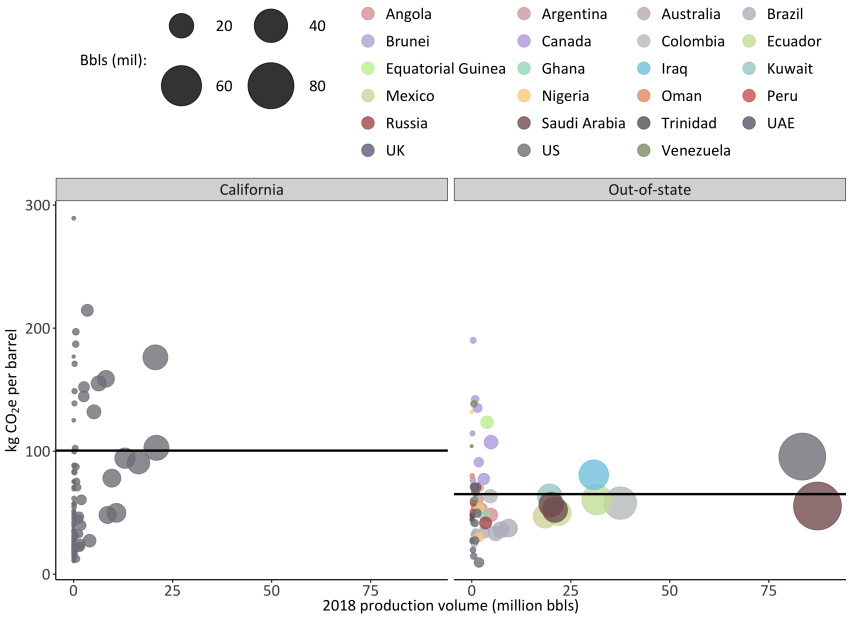
4.1 Field-level extraction greenhouse gas emission factors

To estimate GHG emissions and analyze the impact of macro-economic conditions such as carbon prices on crude oil extraction, both *spatial* heterogeneity and *temporal* variability of GHG emission factors (kg CO_2/bbl) are important. Spatial heterogeneity in GHG factors across oil fields is driven by enhanced oil recovery methods including steam injection for oil extraction amongst other factors (El-Houjeiri et al., 2017). Additionally, changes in GHG emission factors over time within fields are driven by changes in production processes e.g. increase in steam injection driven by a decrease in the price of natural gas relative to that of crude oil.

There are a number of sources to derive GHG emission factors for oil extraction from: CARB's GHG Inventory (California Air Resources Board, n.d.*a*), CARB's Mandatory Greenhouse Gas Reporting Regulation (MRR) program (California Air Resources Board , CARB), and CARB's LCFS Crude Oil Life Cycle Assessment (California Air Resources Board, n.d.*b*). The GHG Inventory provides yearly reported values of GHG emissions (which can be divided by reported oil production in the DOC WellStar dataset to calculate emissions factors), but the GHG Inventory is reported at the state level and therefore does not have any spatial heterogeneity. The MRR dataset provides reported emissions annually at the firm-level, thus providing additionally spatial specification. However, because oil producers/firms operate in multiple fields and a field can have multiple firms operating within it, there is no direct way to obtain field-level emission factors. CARB's LCFS Crude Oil Life Cycle Assessment has the finest spatial resolution for GHG emission factors, providing carbon intensities (gCO_{2eq}/MJ) for 157 oil fields in California. Figure 10 shows the 157 LCFS California emission factors on the left panel, while the right panel are all other non-California emission factors included in the LCFS assessment. The limitation of the LCFS dataset, however, is that the GHG emission factors developed in the assessment are available only for the year 2015. We acknowledge that oil extraction practices, particularly the use of steam injection, may change

from year to year. However, given the importance of incorporating spatial heterogeneity to quantifying GHG emissions in order to understand the impact of carbon prices on field-level crude production, we chose to use CARB's LCFS Crude Oil Life Cycle Assessment as the source of GHG emission factors.

Figure 10: Extraction greenhouse gas emission factors for fields within and outside of California, as derived by the LCFS Assessment.



Data: California Air Resources Board LCFS Crude Oil Life Cycle Assessment (California Air Resources Board, n.d.*b*)

CARB utilized the OPGEE model (El-Houjeiri et al., 2017) to develop the field-level GHG factors that account for the production, processing and transport of crude petroleum. The lifecycle assessment was based on DOC 2015 data on oil production, steam injection, water injection and other field-level extraction practices. Because the CARB assessment provided lifecycle emission factors and our analysis is limited to GHG emissions from the extraction segment only, we re-calculated the GHG emission factors using the OPGEE model and inputs from CARB but limiting upstream emissions to those from exploration, drilling and crude production.

CARB's LCFS Crude Oil Life Cycle Assessment GHG factors are available for only 157 California oil fields or 35% of California's total oil fields. To estimate GHG emission factors for those fields not included in the LCFS Assessment, we first split the fields included in the LCFS Assessment into two groups — fields that utilize steam injection and those that do not — and separately estimate their median upstream GHG emission factors. Using DOC WellStar data (Department of Conservation, n.d.*b*), we then determine whether the excluded

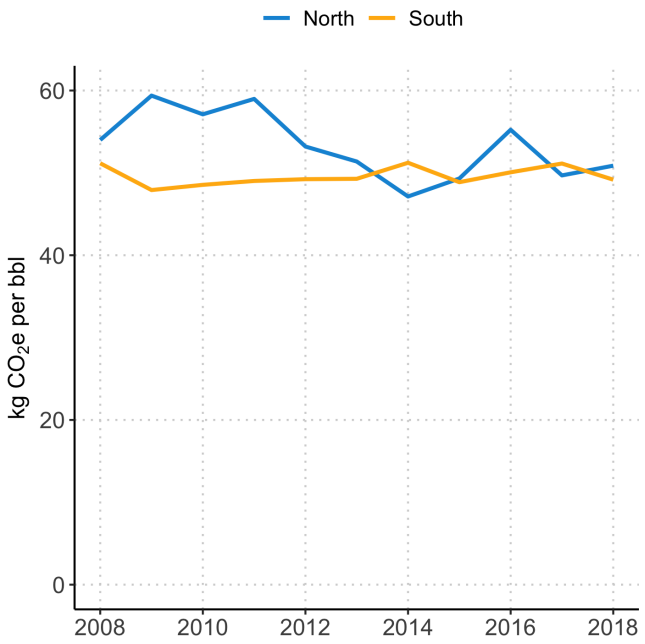
oil fields utilized steam injection in 2015. Based on whether the excluded oil fields utilized steam injection or not, we assign the corresponding median upstream GHG emission factor of fields included in the LCFS Assessment.

Our modeled total GHG emissions for past years differ from those reported in the GHG Inventory because of differences in the two methodologies. Further, although extraction practices including steam and water injection are likely to experience interannual variation within fields, the LCFS GHG emission factors are based on 2015 only. The 2019 baseline for GHG emissions from the extraction segment used in this study is estimated using historical crude extraction from the WellSTAR monthly extraction data (Department of Conservation, n.d.*c*) and the estimated field-level GHG emission factors described previously.

4.2 Cluster-level refining greenhouse gas emission factors

To calculate refining GHG emission factors, we use the CARB MRR dataset's reported annual emissions at individual refineries (California Air Resources Board , CARB). However, because crude consumption data is reported at the cluster level, we are unable to decipher how much crude is consumed at the individual refineries. Therefore, we are unable to calculate refinery-level GHG emission factors. Thus, we aggregate the GHG emissions reported in the MRR dataset to the cluster level (North and South) based on each refinery's location. We then divide the cluster-level GHG emissions by the cluster-level crude input in 2018 as reported by the CEC Weekly Fuels Watch dataset (California Energy Commission, 2020) to estimate cluster-level GHG emission factors. Because of this data limitation, we assume that all refineries within a cluster have the same GHG emission factor. Figure 11 shows how the cluster-level refining GHG emission factors have changed from 2008 to 2018. For our modeled GHG emissions, we use the most recent year of MRR data available (2018).

Figure 11: Cluster-level refining greenhouse gas emission factors over time.



Data: California Air Resources Board (California Air Resources Board , CARB)

4.3 Calculating projected GHG emissions

Within each year for each modeled scenario, we estimate the total amount of GHGs emitted by both, the extraction and refining segments. For the extraction segment, we calculate GHG emissions at the field level by multiplying field-specific GHG emission factors (as derived in Section 4.1) by the amount of oil extracted within a field during that scenario year. The total GHG emissions from oil extraction in a field f in year t, $GHG_{f,t}$, can be calculated as shown in Equation 13.

$$GHG_{f,t} = EF_{ext,f} \times q_{f,t} \tag{13}$$

Where $EF_{ext,f}$ is the GHG emission factor associated with extraction in field f and $q_{f,t}$ is the amount of oil produced in field f in year t.

Similarly, for the refining segment, we estimate the refinery-level GHG emissions by multiplying the GHG emission factors (derived at the cluster level) by the crude oil consumed by that refinery. The total amount of GHG emitted in refining cluster c in year t, $GHG_{c,t}$, can be calculated using Equation 14.

$$GHG_{c,t} = EF_{ref,c} \times V_{c,t} \tag{14}$$

Where $EF_{ref,c}$ is the GHG emission factor associated with refining in cluster c and $V_{c,t}$ is the amount of crude consumed at refineries located in cluster c in year t.

5 Quantifying health impacts from ambient PM2.5 and TACs

5.1 Overview of Approach

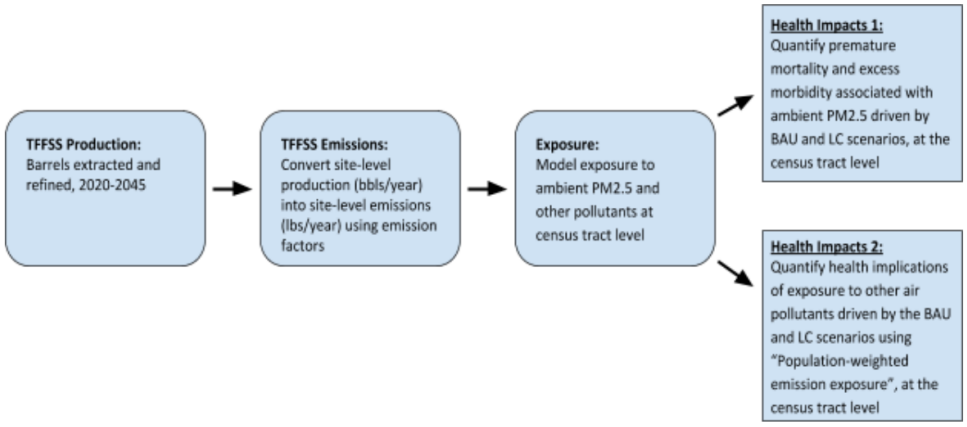
The goal of the health impact analysis is to project the health outcomes estimated under the business-as-usual (BAU) scenario for the operations of the TFFSS from 2019 to 2045 and to contrast to the health outcomes estimated under the low-carbon TFFSS scenarios (LCE1, LCE2, LCR1, LCR2) from 2019 to 2045. The difference in the estimated health outcomes between a low-carbon TFFSS scenario and the BAU scenario is interpreted as the net health impact of the low-carbon scenario.

The BAU and low carbon scenarios are the first input in the health impact analysis. Given those scenarios, the next step is to derive the projected level of production associated with the BAU and LC scenarios. This is done using the supply-side production model described in Chapter 2 to each BAU and low-carbon scenario trajectory, from 2019 to 2045. From the application of the model we obtain a projected level of extraction in bbls per year at the site level for 262 extraction sites in California, and a predicted level of refining in bbls per year at the site level for the 16 refineries sites in California.

Given those scenarios, Figure 12 below provides a stylized illustration of the conceptual framework. The next step is to derive the projected level of production associated with the BAU and LC scenarios. This is done using the supply-side production model described in chapters 2 and 3 to each BAU and low-carbon scenario trajectory, from 2019 to 2045. From the application of the model we obtain a projected level of extraction in bbls per year at the site level for 262 extraction sites in California, and a predicted level of refining in bbls per year at the site level for the 16 refineries sites in California.

These levels of production at the site level and by year from 2019 to 2045 are the main inputs in the health impact analysis. The next steps are performed in this Chapter: (i) converting production into emissions; (ii) modelling exposure to ambient PM2.5 and other pollutants driven by TFFSS emissions; and (iii) quantifying the health impacts of these exposures are explained in the following subsections.

Figure 12: Conceptual Framework for Health Impact Analysis



5.2 Emissions factors

In order to model how changes in crude oil extraction and refining driven by the BAU and TFFSS low-carbon scenarios affect emissions of criteria air pollutants and TACs, we use estimates of emissions factors — the amount of pollutant emitted per barrel of crude oil extracted or refined — for crude oil extraction and refining. We converted production for the business as usual, and scenarios LE1, LE2, LR1, and LR2 by multiplying total barrels extracted or refined per year by emission factors. For air criteria pollutant emissions factors, we rely on published scientific estimates for refineries and extraction activities (Jaramillo and Muller, 2016), which are reported in Table 2. To the best of our knowledge, there are no published emissions factors for TACs in the scientific literature. Therefore, we use data from CEIDARS facilities matched with extraction fields and refineries and calculate the emissions per barrel extracted or refined using data for the 2008-2016 period. The estimated emission factors for TACs are reported in Table 3.

Table 2: Emission Factors Reported in Jaramillo and Muller (2016)

	Refining (kgs/bbl)	**Extraction (kgs/bbl)**
NOx	0.0136	0.0418
PM2.5	0.0036	0.0014
SOx	0.0079	0.0124
VOCs	0.0113	0.0237

Table 3: Estimated Emissions Factors from the CEIDARS Database

	Refining (kgs/bbl)	**Extraction (kgs/bbl)**
Total TACs	0.00085537	0.00394619
Top 10 toxics	0.00003738	0.0006263
1,3-Butadiene	0.0000009	0.00000269
Acrolein	0.00000099	0.00000951
Arsenic	0.00000003	0.0000004
Benzene	0.00001359	0.00009095
Cadmium	0.00000006	0.00000017
Chromium	0.00000001	0.00000001
Formaldehyde	0.00002049	0.00052138
Nickel	0.0000006	0.00000036

5.3 Modelling Exposure to Ambient $PM_{2.5}$

The health impact analysis focuses first on the primary emissions of PM2.5 along with the secondary emissions from precursor criteria pollutants (NOx, SOx, NH3, and VOCs) from the TFFSS under the BAU and low-carbon scenarios. We focus on the health impacts of ambient exposure to PM2.5 because of its well-documented negative effects on health, such as premature mortality, increased incidence of heart and lung problems, and severe acute effects such as increased risk of heart attack and congestive heart failure. Moreover, the main long-term human health effects from NOx, SOx, and VOCs emissions operate through their role as precursors of secondary PM2.5 and ozone. For these reasons, as explained in the Synthesis Report, the majority of air quality impact evaluations performed by leading agencies such as the U.S. EPA and CARB focus on PM2.5.

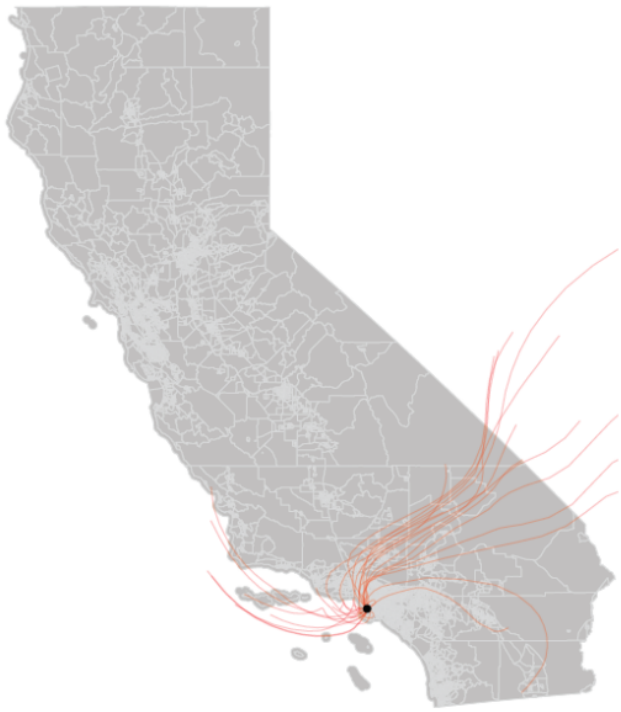
We model the TFFSS emissions of PM2.5 associated with a given BAU or TFFSS LC scenario and convert them into exposure to ambient PM2.5 across California at the census tract level by using an atmospheric transport model that calculates both the primary and secondary release of PM2.5. Specifically, we use the InMAP (Intervention Model for Air Pollution) model, which produces estimates of annual-average changes in primary and secondary PM2.5.

After obtaining emissions from all the pollutants, we linked the refineries emissions to other information regarding refineries such as location and stack characteristics (height, diameter, velocity, and temperature) obtained by CARB. We linked the fields' emissions to the location characteristics of each field. This information is then inputted in InMap and we obtained the $PM_{2.5}$ pollution exposure at the census tract level. InMAP uses physical and chemical information from the output of a state-of-the-science chemical transport model (Tessum, Hill and Marshall, 2017). Furthermore, InMap uses several variables to calcu-

late pollution exposure such as annual averages of wind vectors, wind speed, temperature, pressure, friction velocity, boundary layer height, dry and wet deposition rates, gas/particle phase partitioning of pollutants, and parameters relevant to the calculation of emissions plume rise. Figure 13 illustrates the transport of pollutants in the atmosphere from one refinery in Los Angeles county.

The output from InMap is a grid of ambient concentrations of primary and secondary PM2.5. InMap has two options for calculating ambient concentrations of PM2.5: a uniform 1.5 km-grid resolution and a varying 1km to 48km-grid resolution, where the lowest resolution is associated with urban areas according to Tessum, Hill and Marshall (2017). We used the grid resolution of 1km to 48km given its faster computational time. We then linked the grid cells provided in InMap to census tracts in California by spatially merging InMap's output to California's census tracts. In the cases where one census tract contains more than one grid, we assign the mean of the grids that correspond to the census tract.

Figure 13: Example of Wind Transport Trajectories for a Los Angeles Refinery



5.4 Quantification of health impacts associated with ambient $PM_{2.5}$

The approach to quantify the health impacts from changes in exposure to ambient PM2.5 associated with the BAU and the TFFSS LC scenarios closely follows the methodology in the U.S. EPA BenMAP software (Sacks et al., 2018). This approach is the standard to evaluate the health impacts associated with changes in concentrations of ambient air pollutants and used by the California Air Resources Board (2019*b*). We re-code the BenMAP software in the programming language R to allow for an analysis at the census tract level, which is not readily available in BenMAP. This allows us to contrast health impacts of the scenarios at the community level, in particular comparing impacts between disadvantaged and non-disadvantaged communities. The two categories of health impacts from changes in ambient PM2.5 exposure that are modeled are mortality effects and morbidity effects.

We proceed in two steps: first by calculating the change in health (mortality or morbidity) associated with a change in ambient PM2.5 concentrations, and second by converting the changes in health outcomes in monetary values using the value of statistical life approach or the willingness to pay approach (when applicable).

Step 1: Estimating health impacts at the census-tract level. For a census tract j, age-group a and year t we calculate the change in health associated with the change ambient concentrations of PM2.5 under a given BAU or low-carbon scenario for year t as:

$$\Delta health_{jat} = population_{jat} \times \Delta rate_{jat}(BaselineRate_{jat}, \beta_a, PM_{2.5_{jt}})$$

There are 5 components to this equation:

- Population: This term represents an estimate of the number of people affected by change in ambient PM2.5 concentrations. In practice, population_{iat} is the number of individuals in census tract j, in age group a from the 2010 decennial census and projected to future time period t using age-group specific population growth rates at the county level from California Department of Finance population projections

- Δrate: This term represents a function that estimates of the percentage change in the risk of an health effect due to a one unit change in ambient air pollution. Depending on the epidemiological study, the function can take a linear, log-linear, or logistical form.[9] The function depends in turn on the baseline incidence rate, the β parameter, and the change in ambient PM2.5 concentrations.

- Health baseline incidence: The health baseline incidence rate is an estimate of the number of people who die prematurely (or suffer from some adverse health effect) in a given population over a year, under the baseline level of ambient PM2.5 concentrations.

We obtain baseline incidence data from BenMAP. For mortality incidence between 2019 to 2045, BenMAP uses both the U.S. Centers for Disease Control (CDC) WONDER database and the Census Bureau national mortality rate tables. Morbidity incidence rates are from the Healthcare Cost and Utilization Project (HCUP). Baseline morbidity rates from HCUP are for 2014.

- The coefficient a represents the percentage change in a health outcome associated with a 1 μg/m3 increase in PM2.5 exposure for age group a. This parameter summarizes the concentration-response relationship between PM2.5 and health outcomes. These coefficients come from epidemiology studies used by CARB when studying the health effects of PM2.5, and by the U.S. EPA's Regulatory Impact Analysis (RIA) for the National Ambient Air Quality Standards for PM California Air Resources Board (2019b); **?**. These coefficients are assumed to be the same for all census tract and time periods following our review of the scientific literature.

- PM2.5 represents the change in ambient concentrations of PM2.5 under a BAU or low-carbon scenario relative to a baseline with null ambient PM2.5 concentrations. Changes in ambient PM2.5 are driven by changes in direct emissions of PM2.5 and by changes in the emission of precursors (NOx, SOx, NH3, and VOCs). We obtain the change in ambient PM2.5 from the application of the InMap model and should be interpreted as the difference between ambient PM2.5 concentrations due to productivity activities in the TFSSS in a given scenario minus the (null) level of ambient PM2.5 concentrations from the TFSSS if production activities in the sector were to stop.

The change in health outcome for a given census tract in a given period is the sum of the age-specific impacts for that census tract.

$$\Delta health_{jt} = \sum_{a} \Delta health_{jat}$$

The census tract level calculations will also be done separately for disadvantaged (DAC) and non-disadvantaged census tracts, based on the CalEnviroScreen classification.

Step 2: Monetize change in health outcomes. Following the standard approach in benefit-cost analysis of environmental policies conducted by the U.S. EPA, we use an estimate of the value of statistical life (VSL) to convert the changes in premature mortality associated with a given scenario in monetary values. The VSL does not correspond to the value of an individual life, but rather is derived from estimates of how much people are willing to pay for small reductions in the risk of dying from adverse health conditions caused by air pollution. The values for the VSL follow the U.S. EPA recommendation of $7.4 million ($2006).

The value of avoided morbidity impacts can be thought as having two components: (1) the cost of illness (COI), such as the total medical cost plus the value of lost productivity, and (2) the willingness to pay (WTP) to avoid those health conditions. Because of the absence of WTP estimates for the four morbidity outcomes considered, only COI estimates from the published scientific literature are used (**???**). This approach and the estimates chosen follows the one used in EPA impact analyses that monetized morbidity outcomes. Therefore, the monetized morbidity impacts reported below should be interpreted as lower bound estimates of morbidity costs. Further, COI estimates are assumed to remain constant between 2019 and 2045 in the projections below.

Given an estimate of the VSL or of the valuation of morbidity health benefits, the monetized value of the change in health associated with a BAU or low-carbon scenario at the census tract level is:

$$HealthBenefits_{jt} = \Delta health_{jt} \times MonetaryValue_t$$

Monetized values of the health benefits in future periods are adjusted according to income growth rates using national output projections from the Congressional Budget Office and the income elasticities used in BenMAP. Since the health benefits associated with a BAU and low-carbon scenario will occur in the future (i.e., between 2020 and 2045), we convert future benefit estimates to present time values using a discount factor of 3%.

We now present some information on the inputs used to estimate the health impacts. We obtain population projection data from California's Department of Finance for the period 2020 to 2045. Overall, the state's population is expected to grow by roughly four million over the next two decades and a half. This projected growth in population implies that aggregate exposure to ambient PM2.5 concentrations will grow over time in the state.

Figure **??** further breaks down the overall population projection by age group: less than 18, 18 to 64, and 64 and above. To ease interpretation and display in the figure, we normalize each population series to take a value of 100 in 2019.

The California's Department of Finance projection data indicates that the most of the relative population growth in California will occur in the >64 age group.

Table 4 reports the various estimates of the β_a coefficients, along with the health outcome and population it pertains to and the source published scientific study.

Figure 14: Normalized Population Projection in California by Age Group (2019 to 2045), 2019=100

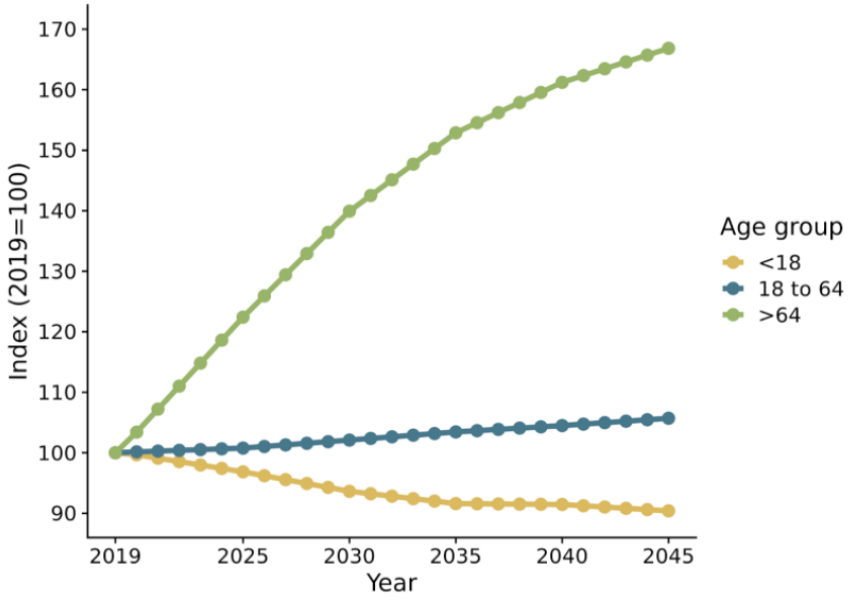


Figure 15: Projected Change in Mortality Rate by Age Group (2019 to 2045), 2019=100

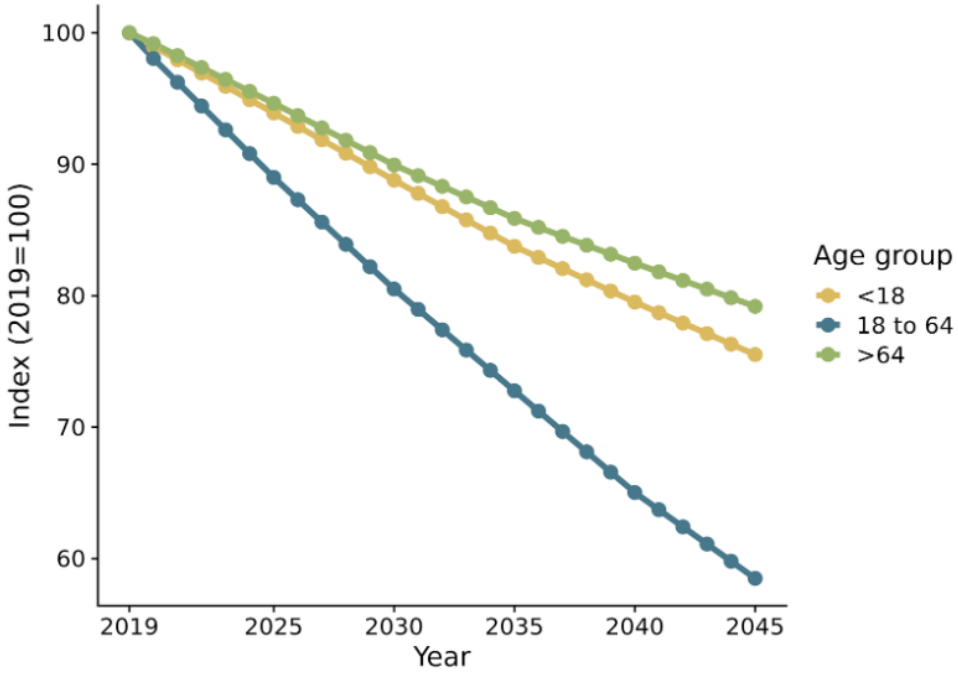


Table 4: β_a Coefficients Measuring the Effects of Exposure to Ambient PM2.5 on Health

Effect (Location, age group a)	**Percent increase in effect from a 1 μg/m3 increase in ambient PM2.5**	**Source**
Premature mortality	0.0058	Krewski et al. (2009)
(US, age >29)	(0.0001)	
Infant mortality	0.0068	Woodruff, Parker and Schoendorf (2006)
(US, age < 1)	(0.0073)	
Cardiovascular hospitalization	0.0019	Zanobetti et al. (2009)
(US, age>64)	(0.0003)	
Respiratory hospitalizations	0.0021	Zanobetti et al. (2009)
(US, age>64)	(0.0004)	
Emergency room visits for asthma	0.0056	Mar, Koenig and Primomo (2010)
(Tacoma, WA, all ages)	(0.0021)	
Non-fatal heart attacks	0.0023	Zanobetti et al. (2009)
(US, age >18)	(0.0006)	

NOTES: The left column lists the key mortality and morbidity health endpoints associated with PM2.5 on an entire population or a subset of population by age group a. The middle column presents the health impact response in percentage points from a 1 ug/m3 increase in air concentration (standard errors are in parentheses). The study from which the estimated parameter was taken is shown in the right column.

5.5 Health Impacts Associated with releases of TACs

As explained in the Synthesis Report, the health impact function approach used to quantify the health impacts associated with exposure to ambient PM2.5 cannot be applied to toxic air contaminants (TACs) since there is insufficient scientific evidence about the β coefficients for TACs.

In order to construct a metric that captures the potential health impacts attributable to exposure of TACs, we construct a measure we call the "population-weighted emissions exposure", or PWEE. The basic principle underlying PWEE is the assumption that the magnitude of health impacts associated with TACs (or other air released pollutants) should reflect the intensity of the release (i.e., the quantity emitted), the population at risk of exposure (i.e. how many people are residing nearby the site where the emission occurred), and the distance to the refining/extraction site.

Therefore we define PWEE for census tract j and year t as the product of the quantity of TAC emitted at a site in census tract j in year t by the population at risk of being exposed. In practice, we assume all the population living within 2 miles from a TAC emitting TFFSS site is at risk of being exposed. The 2 miles cutoff was determined based on the evidence reported in published scientific studies Currie et al. (2015).

We restricted the emission factors to the 10 TACs with highest toxicity given by cancer and non-cancer toxicity weights. In specific, the 10 TACs with highest toxicity are: formaldehyde, asbestos, benzene, nickel, arsenic, cadmium, chromium, acrolein, PAH, and 1,3-Butadiene. To derive the PWEE, we calculated the distance from each TFFSS extrac-

tion and refining site to each census tract within 2 miles based on distance to the census tracts' centroids. We obtained the TACs emitted weighted by distance for each census tract and multiplied this by total population to obtain the PWEE.Equations 15 and 16 shows this procedure. Figure HAF5 illustrates the methodology and results for two TFFSS sites in California: a refinery and an extraction site in Los Angeles county.

$$PWEE_{jrt} = \frac{1}{dist_{jr}} Top10TACs_{rt} \times Population_{jt} \tag{15}$$

$$PWEE_{jft} = \frac{1}{dist_{jf}} Top10TACs_{ft} \times Population_{jt} \tag{16}$$

Where $PWEE$ is the population-weighted emissions exposure at census tract j from extraction field f or refinery site r in year t, $dist$ is the distance from census tract j to extraction field f or refinery site r. $Top10TACs$ are the total highest-toxicity weighted TACs emissions from refinery r or field f in year t. $Population$ is total population in census tract j at year t.

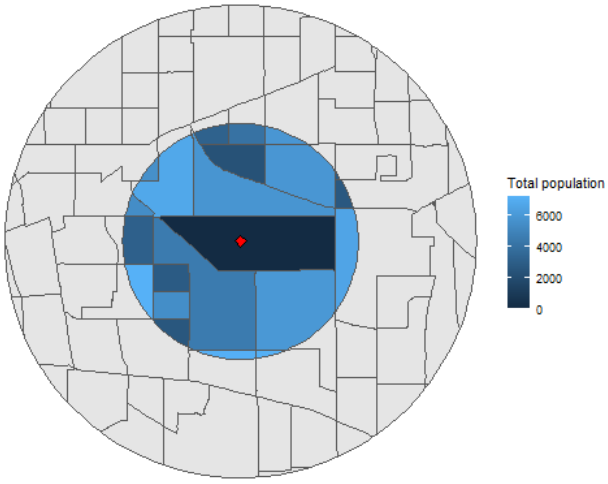
Panels A and B in Figure HF5 show the TFFSS site with the red dot (or area). The boundary lines correspond to census tract boundaries. To compute the PWEE, we then select all census tracts that intersect at any point with a 2 miles radius from the centroid of each TFFSS site. These census tracts with the 2 miles band are shown in blue in the figure, with darker shades indicating larger population in the census tract. Finally, we evaluate the distance between the centroid of each census tract in the blue region to the centroid of the TFFSS site. The PWEE is constructed by multiplying the amount of TACs emitted by each site (in kg) with the population of each census tract in the 2-miles blue region, weighting the census tracts' populations according to the inverse of the distance between the tract and the TFFSS site, so that the populations of census tracts closer to the TFFSS site are given more weight in the calculation.

6 Quantifying labor impacts

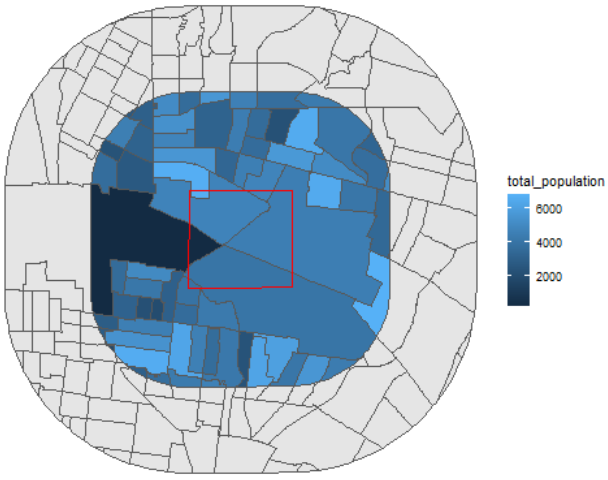
For each scenario in this report we projected barrels of extracted crude oil and refined petroleum produced annually between 2020 and 2045 at the county level. We combine production estimates with the projected annual price of Brent crude oil and refined petroleum products to compute projected annual revenue between 2020 and 2045 separately for the oil and gas extraction and petroleum refining industries. Annual changes in revenue for these

Figure 16: PWEE and Census Tracts within 2 miles of TFFSS Sites

Panel a) Refinery Site in Los Angeles County



Panel b) Extraction Site in Los Angeles County



two industries is then input into IMPLAN (version 5.16), an input-output model which measures the relationship between a given set of demands for final goods and services and the inputs required to satisfy those demands. Government agencies and academics typically use IMPLAN to measure labor market outcomes of policies when empirical data is limited. From IMPLAN we obtain regional measures of employment and worker compensation across California that is associated with a revenue change in extraction or refining. We evaluate the labor market impact of different transportation fossil fuel supply policies by comparing each scenario to a BAU scenario which projects labor market impacts in the absence of regulation.

IMPLAN generates three types of impacts using measures of existing supply chains and spending patterns: Direct, Indirect, and Induced. We define each below.

Direct quantifies the impact on industries primarily affected by a change in economic conditions.

Indirect quantifies the impact of a change in economic conditions on industries connected to primary industries through a supply chain.

Induced quantifies the impact of a change in economic conditions on industries where workers employed by direct and indirect industries spend their income.

We quantify the labor market impact of changes in economic conditions on full-time equivalent job-years and worker compensation. IMPLAN defines full time as an employee working 2,080 hours in 1 year (Clouse (2020*b*)). Furthermore, IMPLAN defines employee compensation as the total payroll cost of a worker (including wages, salaries, benefits, and payroll taxes (Clouse (2020*a*)).

The labor market analysis is split into four parts: baseline jobs and compensation supported by extraction and refining in 2019, projected direct labor market impacts implied by each scenario, projected indirect and induced impacts for each scenario, and the comparison of scenarios.

6.1 Baseline Jobs and Total Compensation Supported by Extraction and Refining in 2019

First, the analysis quantifies the direct, indirect, and induced jobs and compensation currently supported by the extraction and refining industries in 2019. As recommended by IMPLAN, we obtain direct, indirect, and induced impacts using an Industry Contribution Analysis (ICA), which quantifies the jobs and compensation supported by a specified level of revenue in extraction and refining (Lucas (2020)). The direct impacts are computed by inputting the level of revenue observed for the extraction and refining industries separately for each county with active operations. Running this ICA provides a separate refining and

extraction job and compensation impacts for each county with active operations. After downloading employment and compensation impacts for each county by industry, job impacts are converted to full-time equivalents using ratios provided by IMPLAN (Clouse (2020*b*)). Finally, we sum full-time job-year impacts across all industries in a county to obtain the total number of FTE job-years directly supported in each county. Compensation impacts are downloaded from IMPLAN, adjusted from 2020 to 2019 dollars using the Consumer Price Index (CPI), and summed across industries for each county. Table 5 summarizes the input specifications for the direct impact baseline analysis.

Table 5: Input Specifications for Direct Impact Baseline Analysis

Region	Type	Specification	MRIO (Y/N)
Kern	ICA	20 - Oil and gas extraction	N
Santa Barbara	ICA	20 - Oil and gas extraction	N
Ventura	ICA	20 - Oil and gas extraction	N
Orange	ICA	20 - Oil and gas extraction	N
Monterey	ICA	20 - Oil and gas extraction	N
Contra Costa	ICA	20 - Oil and gas extraction	N
Fresno	ICA	20 - Oil and gas extraction	N
Alameda	ICA	20 - Oil and gas extraction	N
Kings	ICA	20 - Oil and gas extraction	N
San Benito	ICA	20 - Oil and gas extraction	N
San Bernardino	ICA	20 - Oil and gas extraction	N
San Luis Obispo	ICA	20 - Oil and gas extraction	N
San Mateo	ICA	20 - Oil and gas extraction	N
Santa Clara	ICA	20 - Oil and gas extraction	N
Tulare	ICA	20 - Oil and gas extraction	N
Kern	ICA	154 - Petroleum refineries	N
Los Angeles	ICA	154 - Petroleum refineries	N
Contra Costa	ICA	154 - Petroleum refineries	N
Solano	ICA	154 - Petroleum refineries	N

While direct impacts quantify jobs and compensation supported in the extraction and refining industries within each county with active operations, indirect and induced impacts include jobs and compensation supported across every county in California. In order to capture the indirect and induced job-years and compensation supported by extraction and refining through supply chains and household spending patterns, this analysis utilizes a Multi-Region Input-Ouput (MRIO) ICA. MRIO quantifies indirect and induced impacts using a model of inter-regional trade, commuting, and household spending patterns (Clouse

(2020*c*)). For more information on MRIO analysis, please refer to this technical paper from IMPLAN (Squibb and Thorvaldson (2020)).

Due to computational complexity, the MRIO analysis is limited to 7 regions (Clouse (2020*c*)), which are chosen by aggregating California's 10 Census regions (State of California (2020)). This analysis combines Census regions 1 and 2, 8 and 9, and 7 and 10. Figure LF1 maps the 7 regions used in MRIO analyses. For extraction, revenue is computed by summing 2019 field-level production to the region level and multiplying by the 2019 IEA price of Brent crude oil as shown in equation 17.

$$R_i^{2019} = p_{2019} \sum_{k=1}^{N} x_{ki}^{2019} \tag{17}$$

For refining, revenue is computed by summing 2019 site-level production to the region level and multiplying by the 2019 IEA price of Brent crude oil plus a product-specific margin capturing the crack spread and retail margins. The product margin is $23 for diesel and gasoline fuels and $20 for jet fuel.

Before inputting the revenue level into IMPLAN's web application, we create the 7 regions by aggregating counties as shown in figure LF1. Extraction and refining revenue is input into IMPLAN for each region that has active operations in 2019. For extraction this includes regions 3, 5, 6, 8 and 9, and 7 and 10. For refining this includes regions 3, 6, and 8 and 9. All inputs are specified as an ICA. Table 6 documents the input specifications for the indirect and induced impact baseline analysis.

Table 6: Input Specifications for Indirect/Induced Impact Baseline Analysis

Region	Type	Specification	MRIO (Y/N)
3	ICA	20 - Oil and gas extraction	Y
5	ICA	20 - Oil and gas extraction	Y
6	ICA	20 - Oil and gas extraction	Y
8 & 9	ICA	20 - Oil and gas extraction	Y
7 & 10	ICA	20 - Oil and gas extraction	Y
3	ICA	154 - Petroleum refineries	Y
6	ICA	154 - Petroleum refineries	Y
8 & 9	ICA	154 - Petroleum refineries	Y

MRIO results are downloaded from IMPLAN's server by region and industry separately for extraction and refining. Employment impacts are converted to FTE job-years using the

IMPLAN ratios mentioned above and compensation impacts are converted to 2019 dollars using the CPI.

Employment and compensation impacts are added across industries to the regional level and apportioned to the county level using employment and compensation shares computed using the 2019 Quarterly Census of Employment and Wages (QCEW). Shares are constructed by computing the share of regional average annual employment and total compensation across all industries, respectively, that is attributable to each county as shown in equation 18. Where x_i is average annual employment or total compensation in county i of region J.

$$share_i^J = \frac{x_i}{\sum_{i=k\in J} x_k} \tag{18}$$

Multiplying regional impacts (z_t^J) by the county-specific shares as shown in equation 19 yields county-level indirect and induced impacts.

$$Impact_{it}^J = z_t^J share_i^J \tag{19}$$

The total amount of FTE job-years and compensation supported by extraction and refining in 2019 statewide are computed by adding direct, indirect, and induced impacts across all counties.

6.2 Projected Direct Labor Market Impacts

To obtain the direct labor market impacts associated with changes in extraction or refining revenue, we run an impact analysis in IMPLAN. While ICA captures job-years and compensation supported by current extraction and refining revenue levels, impact analysis quantifies labor market outcomes associated with changing economic conditions in extraction and refining. Direct impacts ($Direct_{it}$) are computed separately for the 3 extraction and 3 refining scenarios by multiplying annual county-level revenue (in millions of 2019 dollars), R_{it}, by a county-specific employment or compensation multiplier (ω_i) as shown in equation 20.

$$Direct_{it} = R_{it}\omega_i \tag{20}$$

Annual county-level revenue between 2020 and 2045 is computed for extraction (refining) by summing field (site) level production to the county level and multiplying by the IEA price of Brent crude oil in 2019 dollars (price of Brent plus the product-specific margin).

County-specific multipliers are obtained by running an impact analysis for a $1 million change in revenue in each county with projected extraction or refining operations between 2020 and 2045. Job-year multipliers are downloaded from IMPLAN's server, converted to

FTE job-years, and summed across industries to the county level. Compensation multipliers are converted to 2019 dollars using the CPI and summed to the county level.

As shown in equation 21 compensation impacts are reported in present value terms, using a 3 percent discount rate (r).

$$Direct_{it} = \delta^t R_{it} \omega_i, \quad \delta = \frac{1}{1+r} \tag{21}$$

6.3 Projected Indirect and Induced Labor Market Impacts

In order to capture the impact of projected revenue changes through supply chains and spending patterns throughout California, we run an MRIO impact analysis. Due to time constraints, we use IMPLAN to quantify the impacts associated with annual revenue changes for every fifth year between 2020 and 2045 (2025, 2030, 2035, 2040, and 2045) rather than running the model for every year. Annual impacts are then obtained by linearly interpolating the IMPLAN results.

We compute annual extraction (refining) revenue between 2020 and 2045 for each region with active operations by summing field (site) level production to the region level (as defined by the 7 regions in figure LF1) and multiplying by the projected IEA annual price of Brent crude oil (price of Brent crude plus product specific margin).

Since IMPLAN is an annual model and inputs to impact analysis must be in terms of annual revenue changes, we compute the first difference (Δ_t^J) of annual, region-specific revenue separately for extraction and refining as shown in equation 22.

$$\Delta_t^J = R_t^J - R_{t-1}^J, \quad t \in \{2020, 2045\} \tag{22}$$

We run the MRIO impact analysis separately for each scenario (3 for extraction, 3 for refining) and every fifth year between 2020 and 2045. Table 7 displays the input specifications for the indirect and induced impact analysis in IMPLAN. Employment and compensation impacts are downloaded from IMPLAN's server separately for extraction and refining by region and industry. Employment impacts are converted to FTE job-years using IMPLAN ratios and compensation impacts are converted to 2019 dollars using the CPI. We sum impacts for every fifth year between 2020 and 2045 to the county level separately for extraction and refining and linearly interpolate to obtain annual impacts. Since there is a sharp drop in extraction and refining revenue in 2020 that lasts for one year, we assign the annual FTE job-year and compensation impact from 2025 to the years 2020-2024 rather than linearly interpolating.

Table 7: Input Specifications for Projected Indirect/Induced Impact Analysis

Region	Type	Specification	MRIO (Y/N)
3	Industry Output	20 - Oil and gas extraction	Y
5	Industry Output	20 - Oil and gas extraction	Y
6	Industry Output	20 - Oil and gas extraction	Y
8 & 9	Industry Output	20 - Oil and gas extraction	Y
7 & 10	Industry Output	20 - Oil and gas extraction	Y
3	Industry Output	154 - Petroleum refineries	Y
6	Industry Output	154 - Petroleum refineries	Y
8 & 9	Industry Output	154 - Petroleum refineries	Y

Each impact (γ_s^J) represents the annual change in FTE job-years or compensation at the county-level for extraction and refining. We compute the level of FTE job-years or compensation supported by extraction or refining annually between 2020 and 2045 by combining the annual changes with the baseline 2019 levels (z_{2019}^J) as shown in equation 23.

$$Impact_t^J = z_{2019}^J + \sum_{s=2020}^{t} \gamma_s^J \tag{23}$$

Projected annual FTE job-years and compensation are reported in present value terms, using a discount rate of 3 percent and total impacts are computed by summing the direct, indirect, and induced impacts.

6.4 Comparison of Scenarios

We compare two policy scenarios to a baseline separately for extraction and refining. For the statewide comparisons, we calculate cumulative FTE job-years and compensation (in present discounted value terms) by summing the projected annual level of job-years and compensation across all counties and years for each scenario. Then, comparisons are made by subtracting cumulative FTE job-years (compensation) under a policy scenario from FTE job-years (compensation) under the business as usual scenario.

7 Policy levers

This section details the various extraction and refining segment policy levers considered.

7.1 Extraction production quotas

We consider the following four production quota levels:

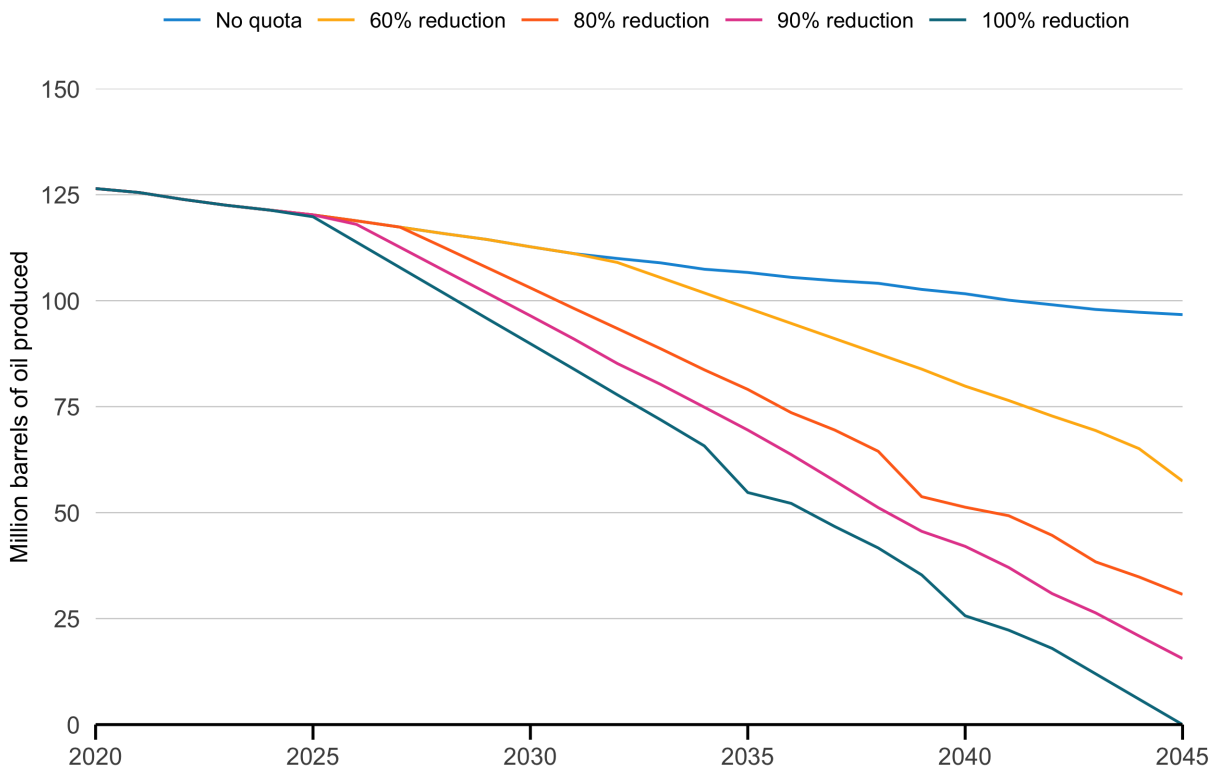
- 60% reduction: annual quota declines linearly from 2019 production (155.7 million barrels) to 40% of 2019 production (62.3 million barrels) in 2045
- 80% reduction: annual quota declines linearly from 2019 production (155.7 million barrels) to 20% of 2019 production (31.1 million barrels) in 2045
- 90% reduction: annual quota declines linearly from 2019 production (155.7 million barrels) to 10% of 2019 production (15.6 million barrels) in 2045
- 100% reduction: annual quota declines linearly from 2019 production (155.7 million barrels) to 0 barrels in 2045

The resulting annual extraction levels are shown in Figure 17. Production quotas are implemented by forecasting field-vintage-level production absent the production quota in time t, and then removing production in a least-cost framework up to the point where the quota is satisfied. The cost for new field-vintages, i.e., those that are predicted to enter from the well entry model for time t, is equal to the sum of field-level capital expenditures (CapEx) and field-level operational costs (OpEx), obtained from Rystad. Each year, OpEx is updated to include carbon prices or CCS, depending on whether the field adopts CCS or not, as discussed Section 8.4. The cost associated with existing field-vintages is OpEx in year t, again adjusted for CCS or carbon prices. Given that existing wells have already incurred capital expenditures, their costs for the purposes of the production quota are only their operational expenditures. Since new wells incur CapEx and existing wells do not, this formulation will generally exclude new wells first to meet the quota, since capital expenditures make new production more costly than production from existing wells. Among existing wells, we assume that older existing vintages will have higher OpEx than newer existing vintages due to relative depletion levels.

If total state-level projected production for a given year exceeds the quota for that year, field-vintages are removed in order of decreasing cost, starting with the most expensive field-vintage, until total production does not violate the quota. In some cases, the last field-vintage removed to meet the quota is a new vintage predicted to enter in time t. In other cases, the last field-vintage removed to meet the quota has already been producing in the historic period. In this case, production from the entire field-vintage is removed. In either case total production may be less than or equal to the actual quota level. The fact that production will rarely exactly equal the quota level corresponds to the discrete nature of new well entry and production in our model. We see this effect illustrated in Figure 17, where production trajectories are not simply straight lines that correspond exactly to the respective quota trajectories discussed above.

Deciding which production to remove in what order can be a feature of the quota policy, and could be tailored to meet alternative policy objectives. The least-cost approach used here is consistent with outcomes that would occur with a production quota implemented by auctioning off a set amount of oil extraction rights. Such a policy would be revenue generating for the state, similar to the State's auction of permits under the GHG cap-and-trade program.

Figure 17: Extraction segment state-wide crude oil production under each production quota



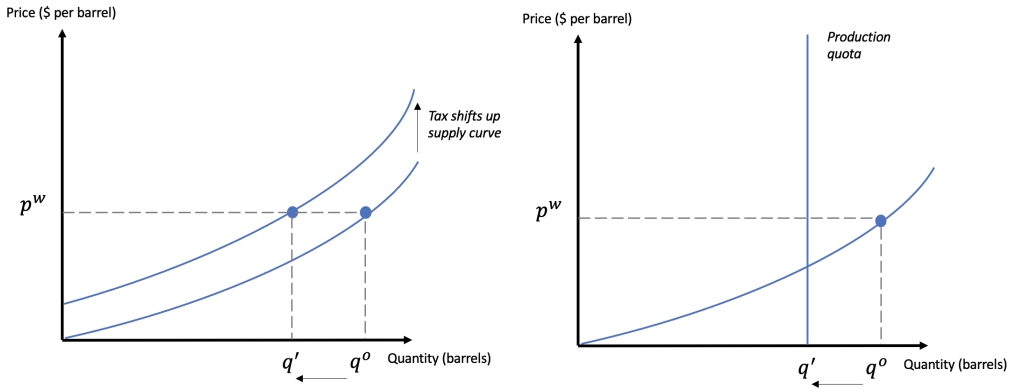
NOTES: This figure shows the four production quotas evaluated as distinct policy levers by the model. All macroeconomic conditions are set to benchmark values.

Production quotas with quota auctioning have similar consequences as under an excise or severance tax on crude oil extraction. Figure 18 shows the correspondence between these two policy instruments. An excise tax can be set to a level that achieves the same reduction in production quantity as in a production quota when the quota is implemented in a least-cost framework such as through auctioning off the right to produce.

The left panel illustrates the impact of an excise tax in this setting. Global oil price p^w determines the total supply of oil extraction from in-state crude oil producers, q^0. A

California excise tax would shift the in-state supply curve of oil extraction up, leading to a reduction of in-state quantity supplied, to q'. The right panel shows the impact of a production quota at q'. We can see that for a given global oil price, there exists a tax and production quota level pair that achieve the same reduction in in-state production quantity.

Figure 18: Correspondence between a production quota and an excise tax



NOTES: The figure illustratively shows the connection between an excise tax and a production quota. The figure demonstrates how both policy levers can be used to achieve the same reduction in crude oil production.

To estimate the tax level that corresponds to each of the evaluated production quotas, we first estimate how in-state production, q^{CA} responds to changes in the global oil prices across our projected scenarios: $\frac{\partial q^{CA}}{\partial p^w}$. Then, we can estimate the excise tax level τ that would lead to the same in-state production quantity q' with the following:

$$\tau = \frac{q^0 - q'}{\frac{\partial q^{CA}}{\partial p^w}}.$$

We can use this expression to estimate a tax level that corresponds to the level of production reduced in the production quota scenarios. The production quotas in Figure 17 show a reduction in million barrels produced compared to BAU, $|q^0 - q'| = \{39, 66, 81, 97\}$ for the increasingly stringent quotas respectively. Under benchmark macroeconomic conditions with no policy, our model results predict that the average change in production per $1 dollar change in oil price, $\frac{\partial q^{CA}}{\partial p^w}$, is equal to 657,517 barrels. Thus, using the expression for τ above, we have that our 60%, 80%, 90% and 100% production quotas could be equivalently achieved with a tax on extraction set at $\{60, 100, 123, 147\}$ $ per barrel (in nominal 2045 USD), respectively.

7.2 Setbacks

We forecast in-state oil extraction quantities under the setback scenarios for both existing wells and new (i.e., projected) wells using the following methodology. To forecast in-state oil extraction quantities under setbacks for existing wells, we first determine which wells would be affected by each of the three setback policy distances. We use GIS analysis to establish circular buffers at three radial distances from sensitive sites. Sensitive sites include homes, schools, healthcare facilities, children daycare facilities, elderly housing, and playgrounds. Data for 3,518 sensitive receptors in California as of July 28, 2020 were provided to this study from the nonprofit organization FracTracker (Ferrar, 2020). The radial distances considered are: 1,000 ft. (0.19 mile), 2,500 ft. (0.47 mile) and 5,280 ft. (1 mile).

Next, for each of the three setback policies, we determine how many wells in each field-vintage (that began producing in the historic period) fall outside of the setback radius, and are thus eligible to produce oil under the setback policy. Note that the spatial data containing coordinates for oil and gas wells does not include all wells in California. Thus, these wells cannot be evaluated in the setback analysis and are assumed to fall outsides of the setback radii considered. Under the "no setback" policy, all existing wells are able to produce in the future. Ineligible wells are removed for the entire duration of the projection period. We then use the model as previously described to forecast the quantity of oil that would be extracted in each field-vintage, given the number of eligible wells, which varies by setback policy.

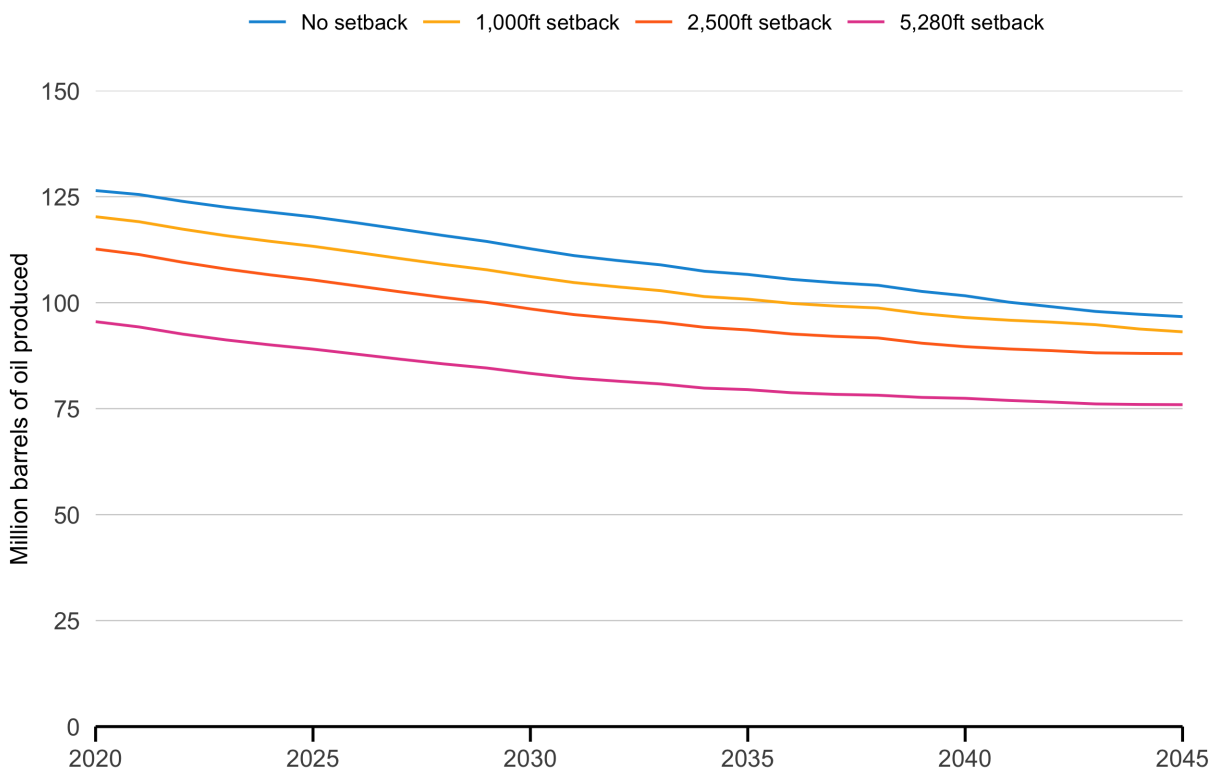
To project in-state oil extraction quantities under setback scenarios for new wells that are forecasted to come online after 2019, the process is slightly different as the model forecasts which field will have new wells, but not the exact location of the new wells within a field. For production from new wells, we first calculate the area in each field that spatially intersects with each setback buffer. For each field-setback policy combination, we determine the area coverage as a percentage of total field area. We use this value to prorate the production quantity that occurs in each new field-vintage under each setback distance. For example, if 70 percent of a field's area intersects with the setback buffer, 70 percent of the production that would have occurred from that field-vintage absent the setback policy is removed under the setback policy. This is modeled by adjusting the peak average well production value for new-well vintages, which is used along with production decline parameters to calculate annual production. If 70 percent of a field intersects with the setback buffer, the peak production value is decreased by 70 percent. The outputs of this setback policy analysis are forecasted oil extraction quantities by field-year.

The method of prorating production from new wells in a setback policy based on area share that intersects with setback buffer was chosen after evaluating several alternative

approaches. First, well entry locations from 1979 to present were analyzed to identify any persistent clustering of well locations. For example, if well locations persistently occurred in one cluster within a field, then the setback analysis should ask whether the setback buffer intersects with that cluster area, and not the entire area of field. However, no persistent clusters areas of well location were identified. Next, well location cluster area trends over time were analyzed. If a time trend were found, for example, such that well locations were moving in persistent directions within fields over time, then these trends could be extrapolated to future years, and the setback analysis could focus on just areas within fields were wells were forecasted to be located. Again, no trend in well locations within fields over time movement were identified. As such, the area prorating method was chosen.

Figure 19 shows the model's estimates of in-state oil production under alternative setback buffer distances. All setback distances lead to reduced in-state oil production in all years relative to no setback policy. As expected, the largest setback buffer, 5280 ft (1 mile), leads to the largest reduction in production compared to no setback policy.

Figure 19: Extraction segment state-wide crude oil production under each setback distance



NOTES: This figure shows state-wide crude oil production under different setback distances. All macroeconomic conditions are set to benchmark values.

7.3 Refining exports

We consider three cases for the export of refined products over the 2020-2045 period.

- **Persistent levels of historical exports:** The average level of exports observed over the period 2014 - 2019 is assumed to continue to 2045.
- **Refineries increase exports to meet historic production standards:** Refineries increase the quantity of exports in order to maintain historic (average of 2015-2019) levels of the capacity utilization factor (CUF) within each refining cluster.
- **Low export scenario:** Exports of gasoline, diesel, and jet fuel linearly decline from 2019 levels to zero in 2045.

The business as usual (no policy) scenario for exports is the persistent historical exports case.

8 Macroeconomic conditions

This section details the various macroeconomic conditions considered during the 2020-2045 forecast period.

8.1 Inflation rate

We assume an annual inflation rate of 2% during the 2020-2045 forecast period. We detailed in each section below where this inflation rate is used.

8.2 Global crude oil prices

We use four projections for the Brent spot price from 2020 to 2045 provided by the US Energy Information Administration (EIA) and the International Energy Agency (IEA) as potential global crude oil price paths in our model. Three oil price paths come from the reference, low, and high oil price trajectories obtained from the EIA's Annual Energy Outlook (AEO) 2020 forecast (Energy Information Administration, 2020*a*). The AEO 2020 oil price forecasts were made before the COVID-19 pandemic. To include an updated oil price trajectory that reflects more current COVID-19 circumstances, we use the projections released in October 2020 under the IEA's World Energy Outlook (WEO) (International Energy Agency, 2020). Specifically, we use the WEO delayed recovery scenario, which projects that global GDP

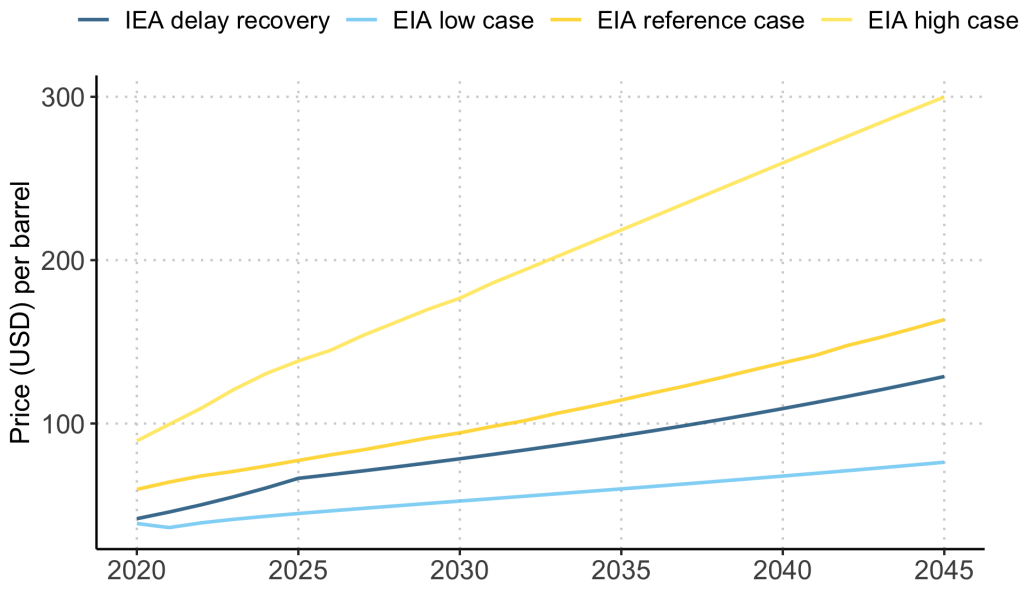
would return to pre-COVID 2019 levels by 2023 as our benchmark oil price scenario as it better reflects updated COVID-19 circumstances.

In summary, our four 2020-2045 global crude oil prices are:

- WEO 2020 delayed recovery oil price (benchmark)
- AEO 2020 reference oil price
- AEO 2020 low oil price
- AEO 2020 high oil price

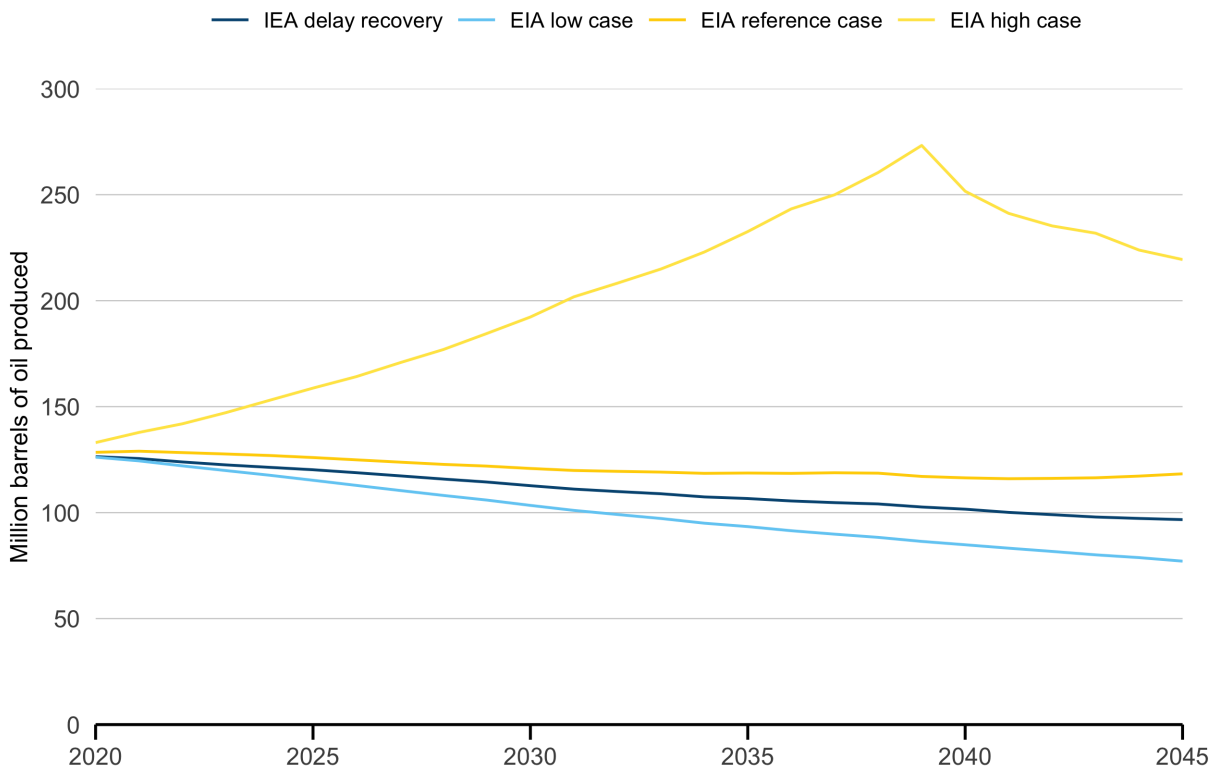
The AEO oil price trajectories are constructed from the annual real price paths provided by the EIA (Energy Information Administration, 2020*a*). The WEO oil price trajectory is constructed from the price forecasts provided for two years in the WEO documentation (International Energy Agency, 2020). Oil price forecasts are given in 2019$ for 2025 and 2040. We calculate the constant real annual growth rates for two periods, between 2019-2025 and between 2025-2040, that would result in these prices. The real annual price growth rate between 2040-2045 is assumed to be the same constant calculated for the period between 2025-2040. The real price for each year is then calculated from these annual growth rates. Finally, we inflate all real prices to nominal prices based on our assumption for future inflation rates (see Section 8.1). Figure 20 shows the projected nominal prices over 2020-2045. Before each model run, an oil price path is chosen. The oil price path enters the well entry model (detailed in Section 2.1) and the CCS adoption choice (detailed below).

Figure 20: Forecasted crude oil price paths (2020-2045)



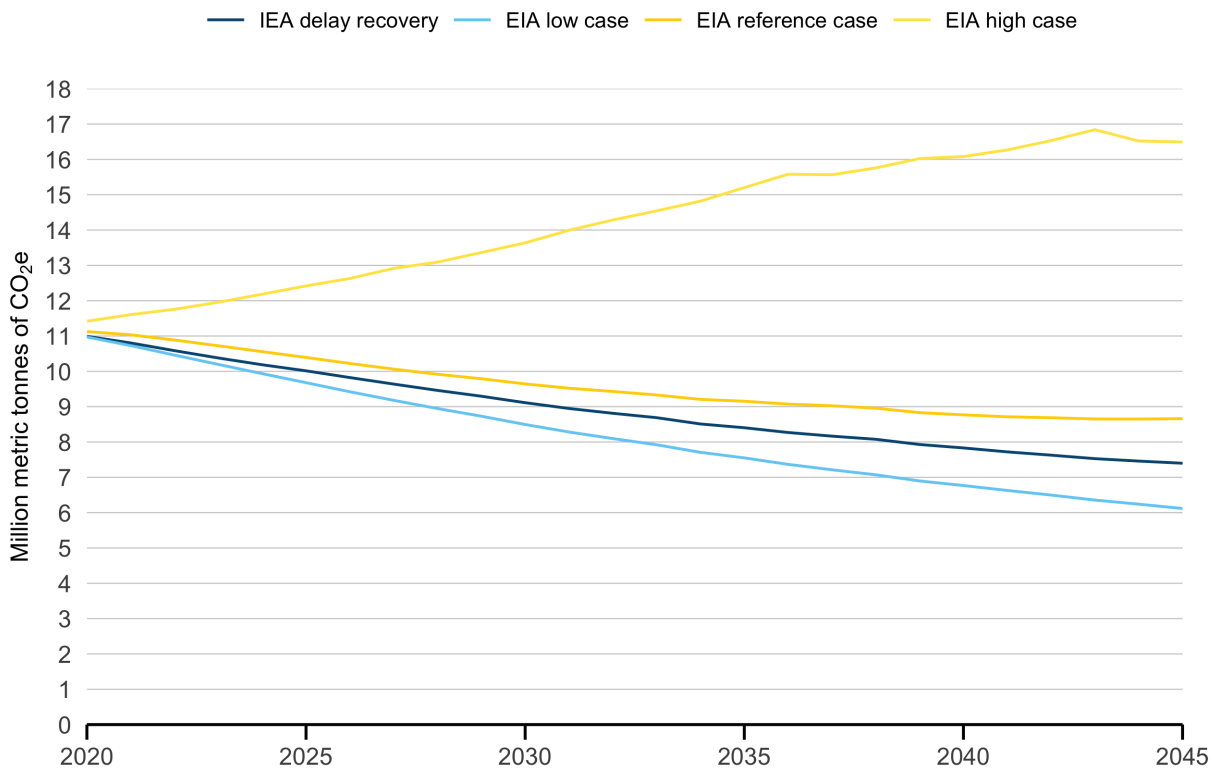
NOTES: This figure plots the four time series of global oil prices, which are used as inputs into the model under four different oil price scenarios. The low, reference, and high price series come from the EIA's 2020 Annual Energy Outlook. The IEA price series comes from the delayed recovery scenario in the 2020 World Energy Outlook.

Figure 21: Statewide crude oil production for each global oil price path



NOTES: This figure shows statewide crude oil production under each global oil price path. All macroeconomic conditions set to benchmark values.

Figure 22: Statewide GHG emissions for each global oil price path



NOTES: This figure shows statewide GHG emissions under each global oil price path. All macroeconomic conditions set to benchmark values.

8.3 California carbon prices

We consider the following three California carbon price paths:

- Low carbon price (a.k.a. price floor): follows California's cap-and-trade allowance price floor, as set by the state's cap-and-trade regulation. The 2020 price floor was announced by the California Air Resources Board in December 2019. (California Air Resources Board, 2019*a*). The prices from 2021-2045 are calculated using a 5% real growth rate described in California Air Resources Board (2019*a*) and our assumed inflation rate of 2% per annum in the future (see Section 8.1).

- Central carbon price (a.k.a central SCC): Follows the central estimate of the social cost of carbon (SCC) estimated by the United States Government Interagency Working Group on Social Cost of Greenhouse Gases (2016). The Interagency Working Group provides SCC estimates in 2007$ in five year intervals: in 2020, 2025, 2030, 2035, 2040,

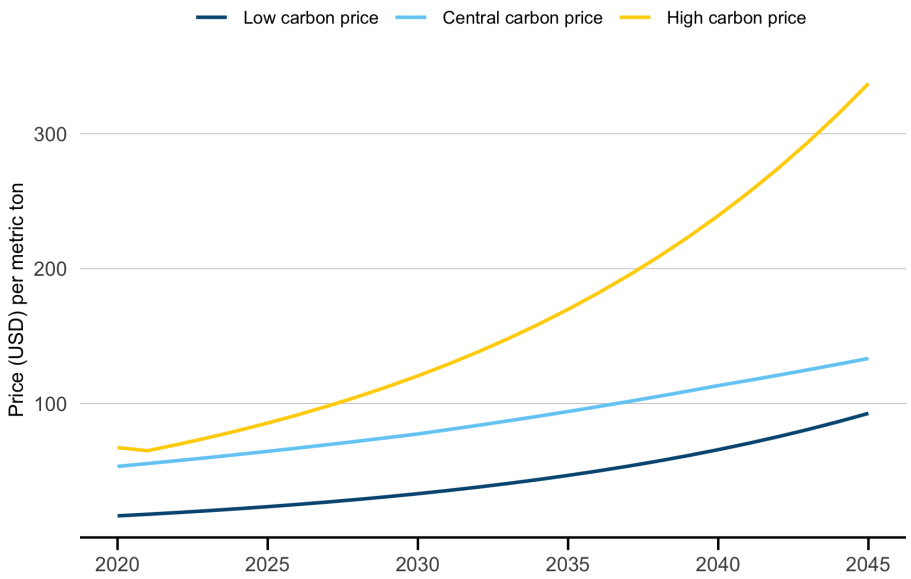
and 2045. We set the central carbon price in these years equal to the SCC values and calculate the carbon price for the years in-between using a linear interpolation. These real carbon prices are then converted to nominal terms using our assumed inflation rate of 2% per annum in the future (see Section 8.1).

- High carbon price (a.k.a. price ceiling): Follows the California's cap-and-trade allowance price ceiling stipulated in Section 95915 in California Air Resources Board (2018). The price ceiling implemented through is $65 in 2021 and rises at a real growth rate of 5%, as prescribed by law, and our assumed inflation rate of 2% per annum in the future (see Section 8.1).

The cost containment reserve in the CT regulation, together with the price ceiling, serves to mitigate high carbon prices by selling allowances at fixed prices. While the reserve tier prices are not used as a carbon price scenario, the reserve prices fall between the low and high carbon price pathways. Thus, our carbon price scenarios encompass carbon price outcomes at the reserve price levels.

Figure 23 shows these paths over time for the projection period.

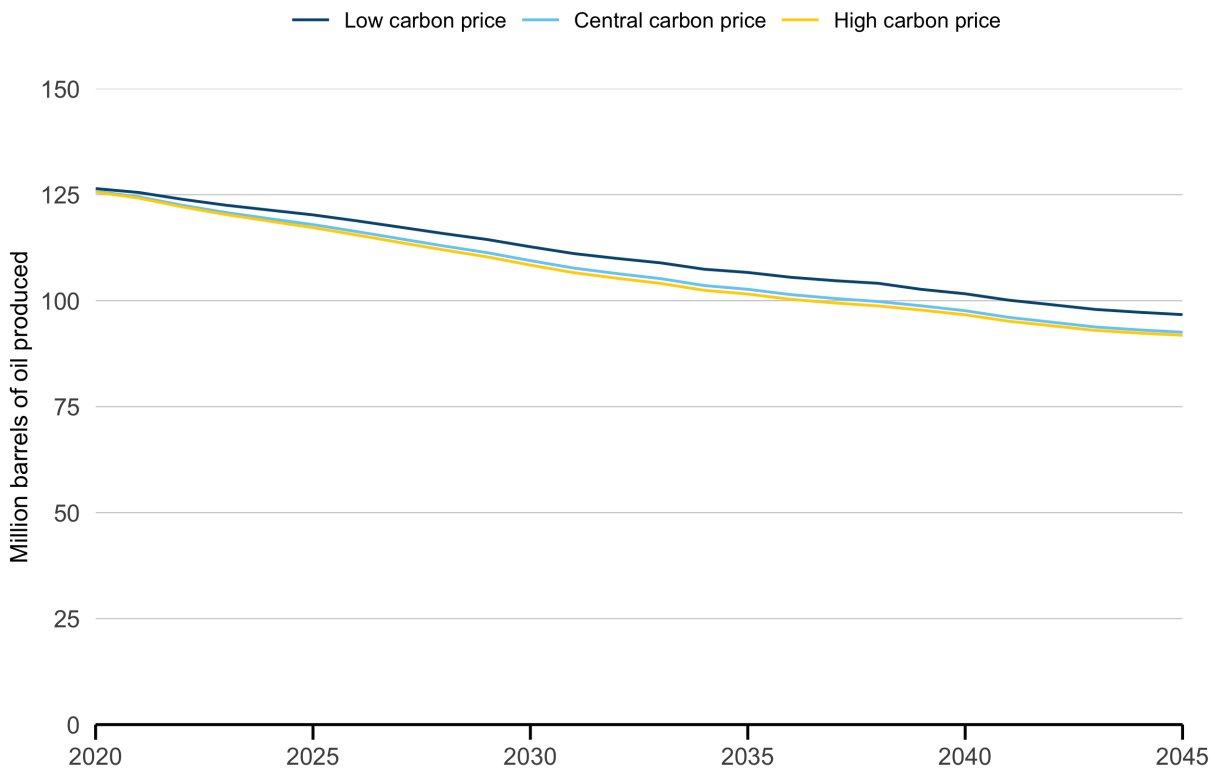
Figure 23: Forecasted carbon price paths (2020-2045)



NOTES: This figure plots the three time series of California carbon prices that are evaluated as different inputs into the model. The lowest and highest series reflect the price floor and ceiling on carbon prices as in California's cap-and-trade regulation (California Air Resources Board, 2018, 2019*a*). The slight dip in the high carbon price from 2020-2021 directly follows from how the regulation specifies the price ceiling's path. The regulation requires specific nominal ceiling values for the years 2012 and 2021, while values for the years between and beyond are to be calculated with a specified growth rate and actual inflation. The 2020 ceiling value (after annual growth and inflation from the 2012 value) ended up higher than the 2021 value. The middle series follows the central estimate of the Interagency Working Group's Social Cost of Carbon (United States Government Interagency Working Group on Social Cost of Greenhouse Gases, 2016).

We select the low carbon price trajectory for our benchmark macroeconomic conditions as the GHG permit prices under California's cap-and-trade program have consistently remained near the price floor since the program's inception.

Figure 24: Statewide crude oil production for each carbon price path



NOTES: This figure shows statewide crude oil production under each CA carbon price path. All macroeconomic conditions set to benchmark values.

8.4 Costs of carbon capture and storage

We evaluate three cases for the current and future costs of carbon capture and storage technologies for oil fields and refineries. Each case represents a time series of prices projected from 2020 to 2045. Our projected prices are developed using cost estimates from the literature as inputs to a model of CCS costs that are a function of the total business as usual emissions. In other words, we treat the literature estimates as averages across oil fields and refineries, where actual costs vary by site as a function of quantity sequestered.

To implement this price projection approach, we first find estimated average costs from the literature. We use the low, central, and high estimated cost of applying CCS to industrial CHP emissions from Benson and Kenderdine (2020), which is: \$75-\$95-\$145 per tCO_2, which includes a capture cost of \$60-\$131, a transportation cost of \$4 per tCO_2, and a storage cost of \$10 per tCO_2. The estimated cost (low-central-high) of applying CCS to refining emissions is \$70-\$80-\$85 per tCO_2, which includes a capture cost of \$58-\$73, a transportation cost of

\$4 per tCO_2, and a storage cost of \$10 per tCO_2 (Benson and Kenderdine, 2020).

Nxt, we model marginal costs with the following:

$$mc_f = b - e_f^{1/\gamma},$$

where e_f denotes the quantity of emissions for a particular oil field or refinery. An expression for total costs can be found by integrating this equation over e; dividing by e provides a measure of average costs. This provides an expression for average costs with two unknowns, b and γ. We set γ to 4 based on our expectation that marginal costs diminish in e. Then, we solve for b using oil field and refinery specific e, based on emissions in 2019. We find the average of the oil field/refinery specific b's for a given literature estimate, and use this value to estimate total costs and average costs by oil field/refinery.

GHG emissions factors for oil fields that adopt CCS are reduced by 87.5% as estimated by Benson and Kenderdine (2020). GHG emissions factors for refineries that adopt CCS are reduced by 58% as estimated by Jing et al. (2020). This occurs in the year in which the oil field and refinery that adopts CCS and for all subsequent years. Chapter 2.g.iii in the accompanying report provides further discussion on these sequestration estimates, the cost estimates, and the associated uncertainties.

Extraction:

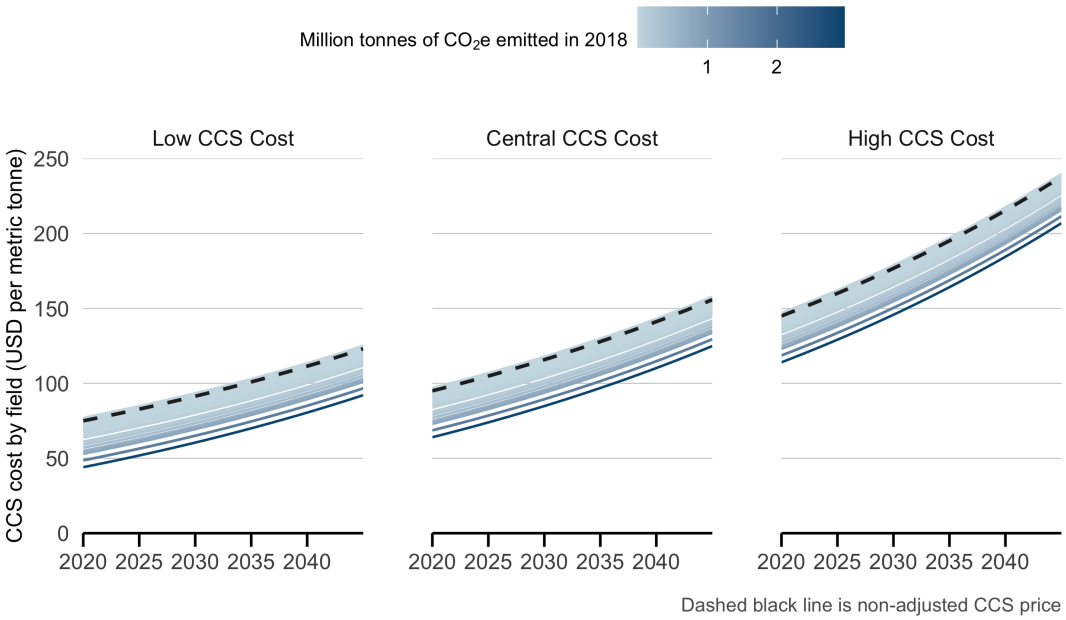
- Low: Constant real cost (in 2020\$) of \$75 per metric ton of CO_2e
- Central: Constant real cost (in 2020\$) of \$95 per metric ton of CO_2e
- High: Constant real cost (in 2020\$) of \$145 per metric ton of CO_2e

Refining:

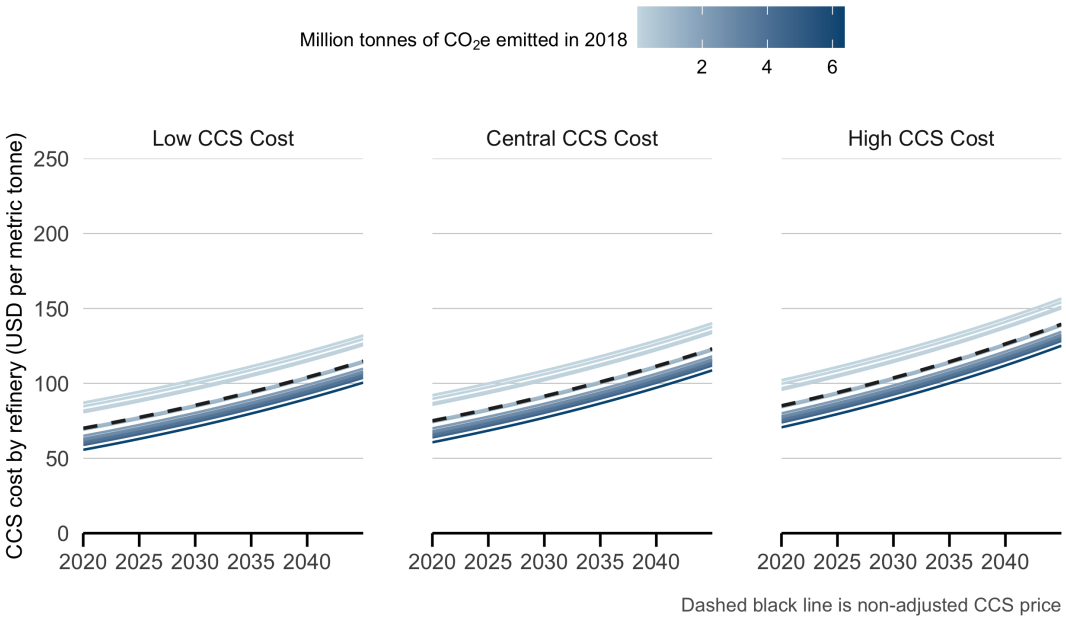
- Low: Constant real cost of (in 2020 \$) of \$70 per metric ton
- Central: Constant real cost (in 2020 \$) of \$75 per metric ton
- High: Constant real cost (in 2020 \$) of \$85 per metric ton

Figure 25a and Figure 25b shows the costs from the literature, as well as the costs by field and refinery after taking the steps above.

Figure 25: CCS costs



(a) CCS costs by oil field



(b) CCS costs by refinery

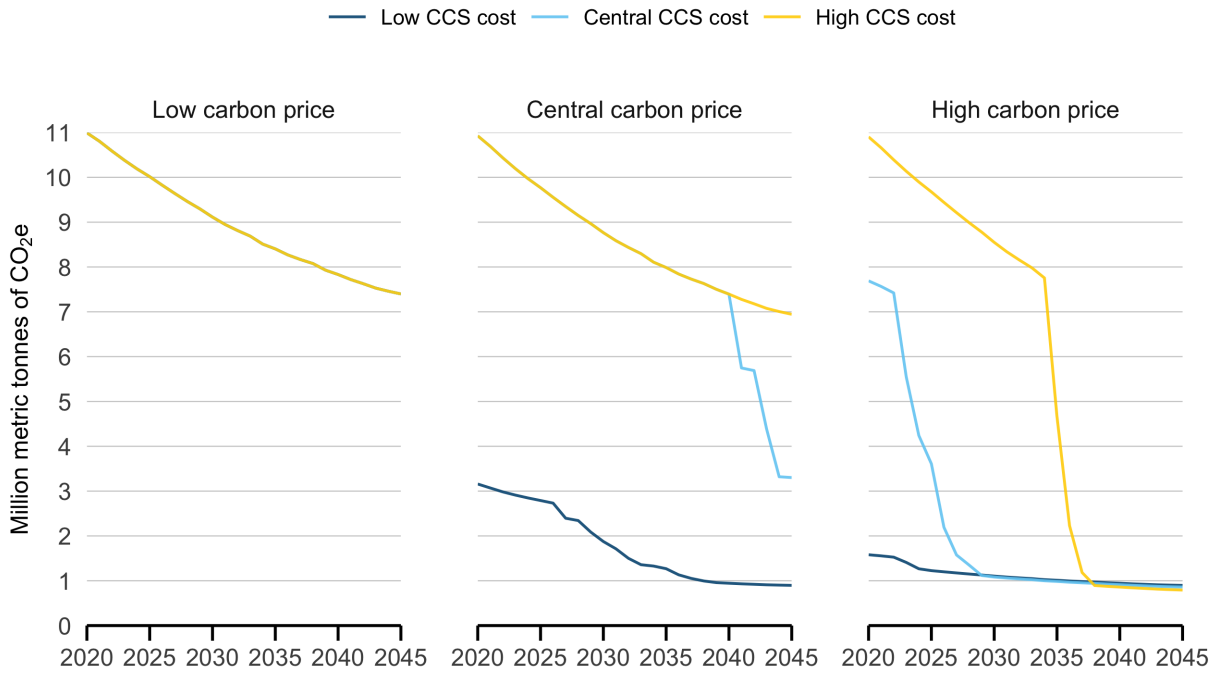
NOTES: These figures plot the CCS costs overall, and by oil field and refinery. The oil field and refinery level costs are estimated by using literature estimates of average costs for each of these sectors, and then estimating oil field and refinery specific costs as a function of quantity of emissions produced by each firm. Details on this procedure are provided in the text above.

GHG emissions reductions from CCS in the model are a function of both CCS costs and carbon prices. Adopting CCS is a method for facilities to avoid paying the carbon price for their emissions (the carbon price is assumed to be a suitable proxy for the price of California cap-and-trade compliance instruments). Thus, these facilities adopt CCS when the cost of CCS are lower than the carbon price per tonne of emissions.

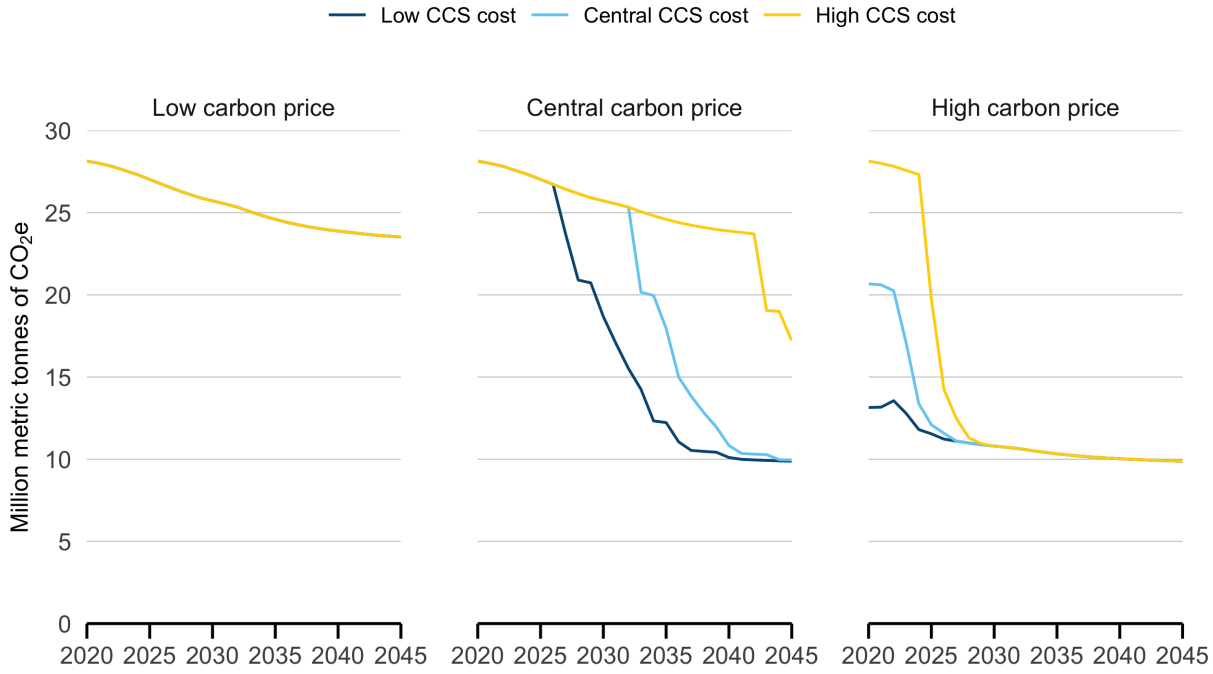
Figure 26a below shows the statewide GHG emissions under alternative carbon price and CCS cost conditions. In general, the figure shows weakly larger reductions in GHG emissions when CCS costs decrease, as well as when carbon prices increase. However, the relationship is not linear. By illustration, the left panel in Figure 26a shows that when carbon prices are low, all CCS costs lead to the same statewide GHG emissions (only one CCS cost time series is shown as the others are overlapping). This is because at very low carbon prices, adopting CCS is never less costly than paying the carbon price. In the central SCC carbon price case, CCS is adopted with low CCS costs, and in some later years, CCS is adopted under central CCS costs. At high CCS costs, CCS is never adopted in the central SCC carbon price case and GHG emissions are the same as in the carbon price floor case. At high carbon prices (carbon price ceiling case), CCS is adopted in some years even under high CCS cost cases. Similar adoption patterns are shown for refineries in Figure 26b.

Figure 26: CCS costs

(a) Statewide GHG emissions by CCS parameter from extraction segment



(b) Statewide GHG emissions by CCS parameter from refining segment



NOTES: The top panel shows statewide GHG emissions from the extraction segment across alternative carbon price cases, by CCS cost case. The lower panel shows statewide GHG emissions from the refining segment across alternative carbon price cases, by CCS cost case.

Benchmark macroeconomic conditions use the central CCS cost estimates from the literature to estimate oil field and refinery-specific costs.

8.5 Innovation

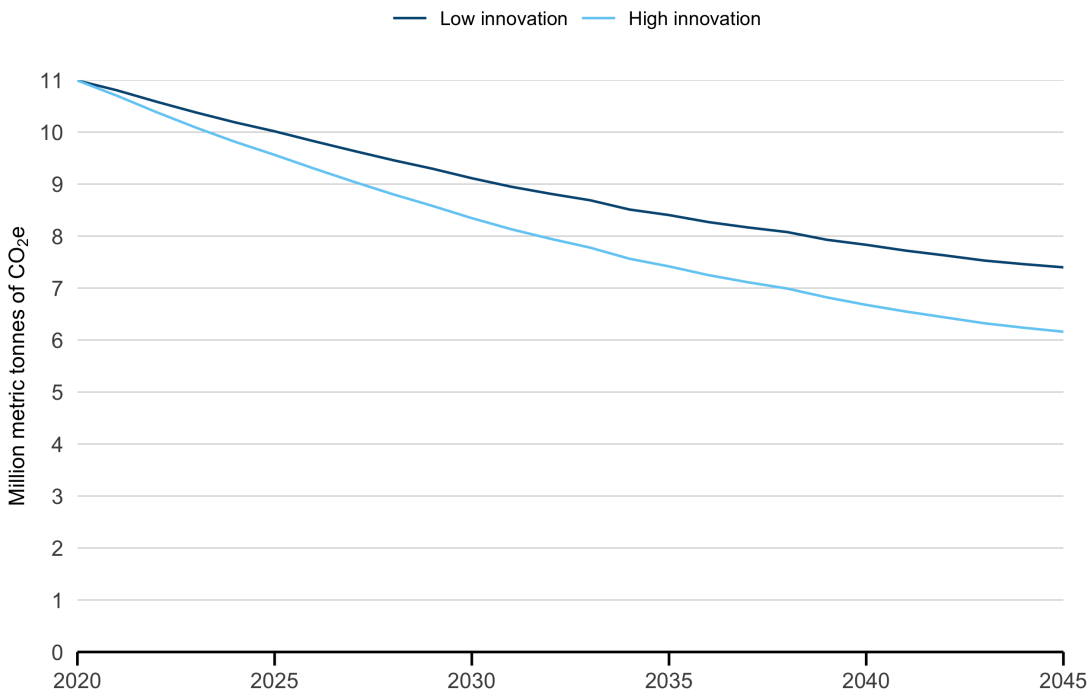
The following two innovation cases are considered:

- **Low innovation:** No improvement in field-level operating costs and GHG emissions factors
- **High innovation:** Operating costs, field-level GHG emissions, and refinery-level GHG emissions factors are reduced by a factor that declines linearly from 1 in 2019 to 0.75 in 2045. This means, for example, that in 2045, each field and refinery GHG emissions factor is 75% of its value in 2019.

To implement these innovation parameters, operational costs and GHG emissions factors are modified before the entry model runs begin. The benchmark cases is that of the "low innovation" case, where GHG emissions factors do not change over the forecast period. This choice of benchmark takes a conservative view of future innovation outcomes.

Figure 27a and Figure 27b below show model's estimates of GHG emissions under Low and High innovation cases. The High innovation case leads to lower GHG emissions in all years for both extraction and refining segments compared to the Low innovation case.

(a) Statewide GHG emissions by innovation case from extraction segment



(b) Statewide GHG emissions by innovation case from refining segment

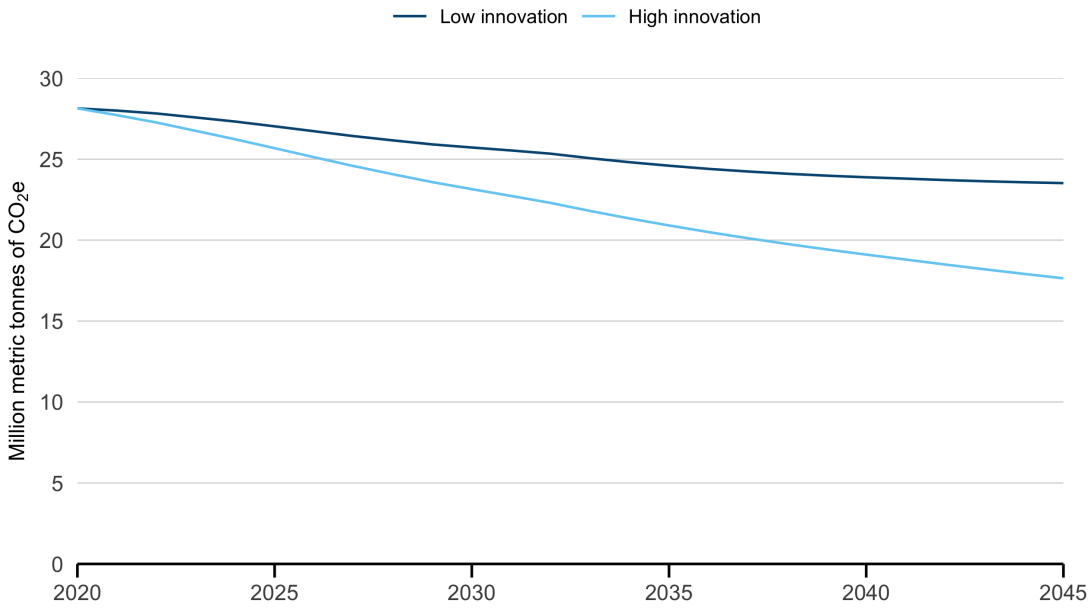


Figure 27: Greenhouse gas emissions under innovation cases

NOTES: The top panel shows statewide GHG emissions by innovation case in the extraction segment. The lower panel shows statewide GHG emissions by innovation case in the refining segment.

9 Scenario selection

This section details our metrics for scenario selection.

9.1 Air exposure benefit

Let t be year between 2019 to 2045. Let f index an oil field (or r in the case of refineries). Let P_{ft} be field-level pollution (in lbs) summed over criteria air pollutants. To account for the number of exposed population and its distribution between disadvantaged communities (DAC) and other communities, we define the state-wide aggregate population-weighted pollution exposure from TFFSS as:

$$\begin{aligned} E_t &= \sum_f P_{ft} w_f^D + \sum_f P_{ft} w_f^N \\ &= E_t^D + E_t^N \end{aligned}$$

where w_f^D is the population (in 2019) within a 2-mile radius of the centroid of field f what is classified as a "disadvantaged community" (i.e., DAC) according to CalEPA's CalEnviron-Screen Score. Likewise, w_f^N is the population (in 2019) within a 2-mile radius of the centroid of field f that is not classified as a non-disadvantaged community (i.e., non-DAC).

Total air exposure benefit We consider two definitions of total air pollution benefit for any particular decarbonization scenario. The first definition is the the percent change in state-wide total population-weighted pollution exposure between 2019 and 2045:

$$\text{Percent change in total pop-weighted air pollution exposure} = \frac{E_{2045} - E_{2019}}{E_{2019}}$$

The second definition is the cumulative state-wide total population-weighted pollution exposure between 2019 and 2045:

$$\text{Cumulative total pop-weighted air pollution exposure} = \sum_{t=2049}^{2045} E_t$$

Distribution of benefit Next, we consider two equity measures to characterize how the total air pollution exposure benefit is distributed between DAC and non-DAC areas. The first measure is simply the difference, or gap, in pollution exposure between DAC and non-DAC areas in year t:

$$\text{Pollution exposure gap}_t = E_t^D - E_t^N$$

Our second equity measure examines what share of the total air pollution exposure benefit is born by DAC areas:

$$\text{DAC share of benefit} = \frac{E^D_{2045} - E^D_{2019}}{E_{2045} - E_{2019}}$$

when DAC share of benefit is 1, all changes in total population-weighted pollution exposure is experienced by DAC areas. Likewise, when that share is 0.5, 50% of the air pollution benefit is borne by non-DAC areas.

9.2 Employment cost

Let R_{it} be revenue (in millions of dollars) from crude oil production for county i in year t.[7]

To account for labor consequences, we define the total employment from TFFSS as:

$$L_t = \sum_i R_{it}\omega_i$$
$$\omega_i = \omega_i^{direct} + \omega_i^{indirect} + \omega_i^{induced}$$

where ω_i is the total employment multiplier from IMPLAN that is applied to revenue from crude oil extraction or refining in county i. The total employment multiplier is the sum of the direct, indirect, and induced employment multipliers for each county in California with active extraction or refining operations. Multipliers are obtained by separately running an impact analysis in IMPLAN for a $1 million change in extraction and refining revenue in each county. Direct, indirect, and induced multipliers by industry are downloaded from IMPLAN's server for each county, converted to full time equivalent job-years using industry-specific ratios provided by IMPLAN, and summed across all industries for each county. Finally, the direct, indirect, and induced multipliers are summed for each county, yielding ω_i.

Total employment cost As with air pollution benefits, we consider two definitions for employment costs for any particular decarbonization scenario. The first is percent change in total employment between 2019 and 2045:

$$\text{Percent change in total employment} = \frac{L_{2045} - L_{2019}}{L_{2019}}$$

[7]To obtain, R_{it}, from field-level output, we first calculate historic production at the field-county level for the most recent year for which production data is available because production in some fields is associated with multiple oil fields. Using this information, we assign future production in each field to counties in proportion to the historical production data.

The second definition is the cumulative state-wide total employment between 2019 and 2045:

$$\text{Cumulative total employment} = \sum_{t=2049}^{2045} L_t$$

9.3 Examining equity and effectiveness differences across policies

Setbacks and job losses Table 8 shows that across oil fields, fields with greater share of its area that would exit production under a setback rule are also in counties with fewer TFFSS employees in 2019. That is, oil fields most affected by setbacks are also those that employ fewer workers. This negative relationship becomes more pronounced as setback distance increases from 1000 ft to 2500 ft to 5280 ft. This suggests that for a given reduction in total air pollution exposure, setbacks of increasing distance lead to fewer job losses. However, we highlight that this mechanism is merely suggestive: none of the regression coefficient shown in Table 8 are statistically significant at conventional levels.

Table 8: Setback-affected oil fields employ fewer workers

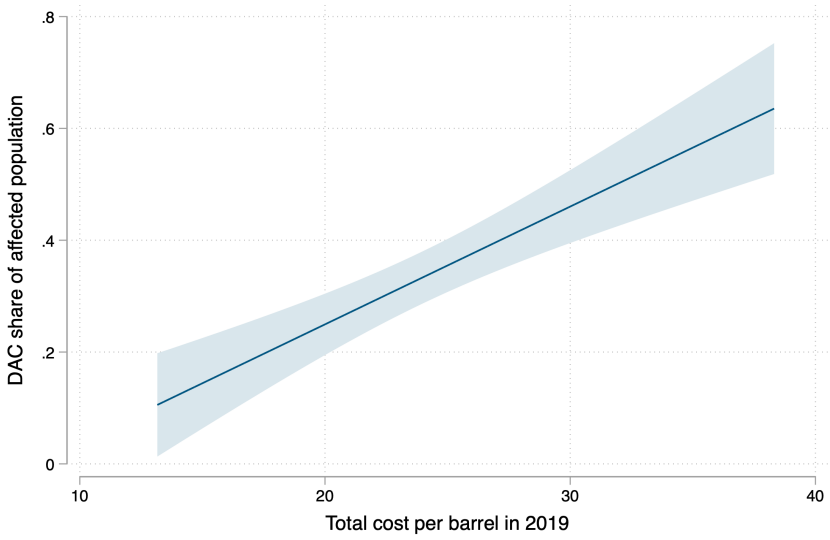
	Outcome is employment		
	(1)	(2)	(3)
Share of oil field affected by setback	-39.52	-52.12	-52.14
	(44.75)	(37.53)	(36.57)
Setback distance	1000ft	2500ft	5280ft
Observations	214	214	214

NOTES: Each column shows the linear coefficient from a separate oil field-level regression model. Outcome variable is TFFSS employment in 2019 associated with an oil field. Regressor is share of oil field area affected by a particular setback distance. Columns 1, 2, and 3 use setback distances of 1000 ft., 2500 ft., and 5280 ft., respectively. Heteroskedastic robust standard errors in parentheses. P-values from two-sided t-tests with *** p<0.01, ** p<0.05, * p<0.1.

Production quotas and equity Why might the introduction of setbacks to a production quota lead to less equity? Figures 28 and 29 offer a potential explanation for why the introduction of setbacks to a production quota may lead to less equity. At the field level, Figure 28 shows a positive correlation between oil extraction cost per barrel and the DAC share of population within a 2 mile buffer of the centroid of the oil field. Under a production quota, oil fields exit from production in the order of high to low extraction costs. This generates an equity-dividend alongside declines in total air pollution exposure: fields that go off line earlier are also fields with air pollution that disproportionately affect more DAC

areas. Thus, DAC areas experience more of the air quality benefit as oil extraction declines under a production quota.

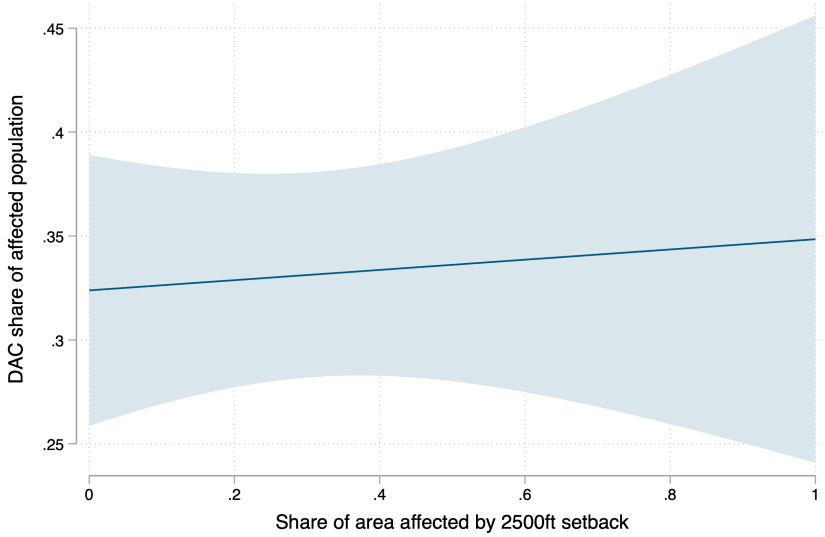
Figure 28: Oil field-level relationship between DAC share of nearby population and cost of extraction



NOTES: Graph shows the oil field-level linear relationship between the DAC share of population within a 2-mile radius of a field's geographical centroid and the sum of operating and capital costs per barrel in 2019. 95% confidence intervals shown.

The addition of setbacks onto production quotas leads to a reordering of which oil fields go out of production. In particular, oil fields that are more closely located to residential and commercial property will tend to exit production earlier. Figure 29 shows a much weaker relationship between an oil field's area affected by a setback - which captures the field's proximity to residential and commercial property - and its DAC share of affected population. Thus, the added presence of setbacks reallocates production cuts to fields that have fewer nearby DAC areas. By having these fields go offline rather than fields that affected more DAC areas, setbacks achieve the same reduction in total pollution exposure but with less equity gain.

Figure 29: Oil field-level relationship between DAC share of nearby population and share of area affected by setback



NOTES: Graph shows the oil field-level linear relationship between the DAC share of population within a 2-mile radius of a field's geographical centroid and the share of field area affected by a 2500 ft. setback. 95% confidence intervals shown.

10 Quantifying property tax impacts

10.1 How property tax is computed in California

Each property tax bill in California contains the assessed value (AV) of land and improvements. Assessed value is the taxable value of the property, which includes the land and any improvements made to the land, such as buildings, well drillings, or other developments. The AV of land and improvements is important because the 1 percent rate and voter–approved debt rates are levied as a percentage of this value, implying that changes in AV result in changes in total property tax collected.

While there are various ways for an assessor to appraise the AV of a land, the "income approach" is used most often by oil and gas (OG) property appraisers due to the unique nature of fossil fuel values. Rule 8 under the mandates of Article XIII A of the California Constitution (Proposition 13) discusses the "income approach" to value wherein a future income stream is converted into a value estimate. Subdivision (b) states that in using the income approach, "an appraiser values an income property by computing the present worth of a future income stream"[8]

[8]Available here: `https://www.boe.ca.gov/proptaxes/pdf/ah566.pdf`

The *Assessors' Handbook Section 566, Assessment of Petroleum Properties* details integrating engineering estimates of the yield curve of each OG producing entities with the latest market forecasts of prices and economic conditions to apply discounted cash flow analysis to income-producing properties such as OG properties.

10.2 Overview of approach

Based upon the preceding section and data limitations, we use a discounted cash flow analysis to compute net changes in property tax revenues from decreased production caused by the policy levers in each scenario (i.e., LCE1: 20% quota, and LCE2: 20% quota + 2,500 ft. setback).

Because we are not modeling new reserves, nor improvements or investments in oil and gas infrastructure upon the land, we are unable to accurately forecast the business-as-usual (BAU) tax revenue base of counties in 2020-2045. Instead we compute the change in tax revenue relative to BAU due to the projected decreased future productions, which are then converted into profits and discounted to the base-year.

Crude oil properties extract less under LCE1 and LCE2 compared to the BAU case, and this difference is due to "legally enforceable government restrictions." Hence, the landowner no longer has the right to remove petroleum and natural gas from the earth "as if it were an unencumbered or unrestricted fee-simple interest" or that the legally "recoverable reserves" has decreased. Based upon the tax code, this deduction in "recoverable reserves" should not be taxed.

It is reasonable to assume that there will be a reassessment or drop in "fair market value" of OG property values when the quota system is implemented. The loss of value from each property is the future missed production profits (denote below as ΔV) due to the quota. As this quota is a "legally enforceable government restriction", we remove this value from the tax base each year.

10.3 Methodology

Let i denote counties $i, \in \{Alameda, ..., Yuba\}$. t denotes years, $t = 0, ..., 25$, where 0 indexes years from 2020. s denotes scenarios, $s \in \{LCE1, LCE2\}$. Next, define the following variables:

- Y_{its}: projected volume in (bbl) of oil produced for county i in year t for scenario s
- P_t : price (\$/bbl) of global crude oil in year t based on the IEA delayed recovery scenario

- γ : average pre-tax operating margin for oil and gas producing entities of 8%
- δ: discount rate of 2%
- T_i: total effective tax rate for county i.[9]

Further define:

- ΔV_{its}: as the change in assessed value for county i in year t due to the cumulative decreased production in scenario s compared to the BAU case.
- ΔR_{its}: as the change in tax revenue for county i in year t due to the decreased production in scenario s due compared to the BAU case.

Using these variables, we construct:

$$\Delta V_{its} = \sum_{t'=t}^{25} \frac{(Y_{it's} - Y_{it',BAU}) \cdot P_t \cdot \gamma}{(1+\delta)^{t'}}$$

ΔV_{its} is the change in tax base for OG property based on the projected cumulative loss in production, $Y_{it's} - Y_{it',BAU}$, due to the scenario s. This future loss in production is converted into US$2020 value term using the 2% discount rate, projected future oil prices (P_t), and industry pre-tax operating margin (γ).
To illustrate, constructing the tax revenue loss in year 2040 (as t indexes year from 2020, $t = 20$) for county i and scenario $LCE1$ would look like:

$$\begin{aligned}\Delta V_{i,20,LCE1} = &\frac{(Y_{i,20,LCE1} - Y_{i,20,BAU}) \cdot P_{20} \cdot \gamma}{(1+\delta)^{20}} \\ &+ \frac{(Y_{i,21,LCE1} - Y_{i,21,BAU}) \cdot P_{21} \cdot \gamma}{(1+\delta)^{21}} + ... \\ &+ \frac{(Y_{i,25,LCE1} - Y_{i,25,BAU}) \cdot P_{25} \cdot \gamma}{(1+\delta)^{25}}\end{aligned}$$

Finally, we apply county ad-valorem property tax to each of these county-year estimate to get the county-year estimates of tax revenue loss. For the time-series plot, we re-scale each year's estimate to the current year's value.

$$\Delta R_{its} = \Delta V_{its} \cdot T_i \cdot (1+\delta)^t$$

[9]This includes the 1% general tax levy from Proposition 13 and voter–approved debt rates that ranges from 0.2% to 0.25%. In our computations, we use the average effective county tax rate observed from 2012-2019 taken from the CA BOE data portal, available: `https://www.boe.ca.gov/dataportal/catalog.htm?category=Property%20Taxes`

ΔR_{its} is the variable that we present in section 4d, "County-level property taxes" of our main report. We also report an annual average tax loss for each county and scenario in million 2020 US$, computed by:

$$\Delta \overline{R_{is}} = \sum_{t=0}^{25} \Delta V_{its} \cdot T_i$$

10.4 Assumptions

Our methodology maintains the following assumptions:

- The assessor has perfect information on when and where the quota will be restricting productions for counties over the period of 2020-2045.
- The tax code will either be flexible enough, or there will be a new proposition passed to take into account the change in market value of OG properties due to the quotas and setback.[10]
- The effective tax rate remains at the average 2012-2019 total county tax rate.[11]
- There are no property conversion in the counties from OG properties to other forms such as commercial, agricultural or residential.[12].

[10]This assumption is reasonable since there are frequent assessment appeals and lawsuits lodged by petroleum companies to lower their assessed value for tax purposes. See here: `https://caselaw.findlaw.com/ca-court-of-appeal/1681888.html`

[11]While there has been slight increases in counties' total tax rates, they have remained fairly stable over the period.

[12]In reality, this assumption is the most likely to be violated as most of the counties have seen assessed value growth from residential, commercial, and agricultural sector while at the same time experienced a drop in the relative tax base share of OG producing properties. See here: `https://www.kerncounty.com/home/showpublisheddocument?id=4797`

References

Anderson, Soren T., Ryan Kellogg, James M. Sallee, and Richard T. Curtin. 2011. "Forecasting Gasoline Prices Using Consumer Surveys." *American Economic Review*, 101(3): 110–114.

Bahrenian, Aniss, Jesse Gage, Sudhakar Konala, Bob McBride, Mark Palmere, Charles Smith, and Ysbrand van der Werf. 2017. "Transportation Energy Demand Forecast, 2018-2030." California Energy Commission CEC-200-2017-010.

Benson, Sally, and Melanie Kenderdine. 2020. "An action plan for carbon capture and storage in California: Opportunities, challenges, and solutions." Energy Futures Initiative, Stanford University.

Borenstein, Severin. 2020. "California's Mystery Gasoline Surcharge Strikes Back." Library Catalog: energyathaas.wordpress.com.

Bureau of Labor Statistics. 2020. "BLS CPI-U."

California Air Resources Board. 2018. "Final Regulation Order: California Cap on Greenhouse Gas Emissions and Market-based Compliance Mechanisms."

California Air Resources Board. 2019*a*. "2020 Annual Auction Reserve Price Notice."

California Air Resources Board. 2019*b*. "Estimating the Health Benefits Associated with Reductions in PM and NOX Emissions."

California Air Resources Board. n.d.*a*. "2000-2018 GHG Inventory (2020 Edition)."

California Air Resources Board. n.d.*b*. "LCFS Crude Oil Life Cycle Assessment."

California Air Resources Board. n.d.*c*. "Low Carbon Fuel Standards (LCFS) Credit Recipients."

California Air Resources Board (CARB). n.d.. "Mandatory GHG Reporting - Reported Emissions."

California Department of Conservation, Geologic Energy Management Division. 2020. "All Wells."

California Energy Commission. 2020. "California Energy Commission Weekly Fuels Watch Report."

California Energy Commission (CEC). 2019. "Finished Products Movements."

Clouse, Candi. 2020*a*. "Employee Compensation."

Clouse, Candi. 2020*b*. "IMPLAN to FTE & Income Conversions."

Clouse, Candi. 2020*c*. "MRIO: Introduction to Multi-Regional Input-Output Analysis."

Currie, Janet, Lucas Davis, Michael Greenstone, and Reed Walker. 2015. "Environmental Health Risks and Housing Values: Evidence from 1,600 Toxic Plant Openings and Closings." *American Economic Review*, 105(2): 678–709.

Department of Conservation. n.d.*a*. "WellSTAR All Wells."

Department of Conservation. n.d.*b*. "WellSTAR Oil and Gas Well Monthly Injection."

Department of Conservation. n.d.*c*. "WellSTAR Oil and Gas Well Monthly Production."

Department of Conservation. n.d.*d*. "WellSTAR Wells 2019."

Department of Conservation Division of Oil, Gas, and Geothermal Resources. n.d.. "California Oil & Gas Fields Volume II - Southern, Central, Coastal, and Offshore California Oil and Gas Fields."

El-Houjeiri, Hassan M, Mohammad S Masnadi, Kourosh Vafi, James Duffy, and Adam R Brandt. 2017. "Oil Production Greenhouse Gas Emissions Estimator OPGEE v2.0." 219.

Energy Information Administration. 2020*a*. "Annual Energy Outlook 2020 - Table: Table 12. Petroleum and Other Liquids Prices."

Energy Information Administration. 2020*b*. "British Thermal Unit Conversion Factors."

Energy Information Administration. 2020*c*. "California Number and Capacity of Petroleum Refineries."

Energy Information Administration. 2020*d*. "EIA Brent."

Energy Information Administration. 2020*e*. "EIA California Refiner Petroleum Product Prices - Sales for Resale."

Energy Information Administration. 2020*f*. "EIA Production Decline Curve Analysis."

ExxonMobil. 2020. "ExxonMobil and Global Clean Energy Holdings Sign Agreement for Renewable Diesel."

Ferrar, Kyle. 2020. "California Setback Analyses Summary."

IMPLAN Group. 2020. "IMPLAN Data Sources."

International Energy Agency. 2020. "World Energy Model Documentation: Table 4: Fossil Fuel Prices by Scenario."

Jaramillo, Paulina, and Nicholas Z. Muller. 2016. "Air pollution emissions and damages from energy production in the U.S.: 2002–2011." *Energy Policy*, 90: 202–211.

Jing, Liang, Hassan M. El-Houjeiri, Jean-Christophe Monfort, Adam R. Brandt, Mohammad S. Masnadi, Deborah Gordon, and Joule A. Bergerson. 2020. "Carbon intensity of global crude oil refining and mitigation potential." *Nature Climate Change*, 1–7.

Kellogg, Ryan. 2014. "The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling." *American Economic Review*, 104(6): 1698–1734.

Krewski, Daniel, Michael Jerrett, Richard Burnett, Renjun Ma, Edward Hughes, Yuanli Shi, Michelle Turner, C Arden Pope III, George Thurston, Eugenia Callee, and Michael Thun. 2009. "Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality."

Lucas, Maria. 2020. "ICA: Introduction to Industry Contribution Analysis."

Mar, Therese F., Jane Q. Koenig, and Janet Primomo. 2010. "Associations between asthma emergency visits and particulate matter sources, including diesel emissions from stationary generators in Tacoma, Washington." *Inhalation Toxicology*, 22(6): 445–448.

Phillips 66. 2020. "Phillips 66 Plans to Transform San Francisco Refinery into World's Largest Renewable Fuels Plant."

Rystad Energy. 2020*a*. "Rystad Asset Wells."

Rystad Energy. 2020*b*. "Rystad Brent."

Rystad Energy. 2020*c*. "Rystad Costs."

Rystad Energy. 2020*d*. "Rystad MyProduction."

Rystad Energy. 2020*e*. "Rystad Production."

Sacks, Jason D., Jennifer M. Lloyd, Yun Zhu, Jim Anderton, Carey J. Jang, Bryan Hubbell, and Neal Fann. 2018. "The Environmental Benefits Mapping and Analysis Program – Community Edition (BenMAP–CE): A tool to estimate the health and economic benefits of reducing air pollution." *Environmental Modelling & Software*, 104: 118–129.

Squibb, Jimmy, and Jenny Thorvaldson. 2020. "IMPLAN's Gravity Model and Trade Flow RPCs." IMPLAN Group Technical Report.

State of California. 2020. "Regions | CA Census."

Tessum, Christopher W., Jason D. Hill, and Julian D. Marshall. 2017. "InMAP: A model for air pollution interventions." *PLOS ONE*, 12(4): e0176131.

Tuttle, Robert. 2020. "Massive Refiners Are Turning into Biofuel Plants in the West." *Bloomberg Green*.

United States Government Interagency Working Group on Social Cost of Greenhouse Gases. 2016. "Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866."

Woodruff, Tracey J., Jennifer D. Parker, and Kenneth C. Schoendorf. 2006. "Fine Particulate Matter (PM2.5) Air Pollution and Selected Causes of Postneonatal Infant Mortality in California." *Environmental Health Perspectives*, 114(5): 786–790.

Zanobetti, Antonella, Meredith Franklin, Petros Koutrakis, and Joel Schwartz. 2009. "Fine particulate air pollution and its components in association with cause-specific emergency admissions." *Environmental Health: A Global Access Science Source*, 8: 58.