Agricultural Council
of California

April 28, 2016

Clerk of the Board
California Air Resources Board
1001 I Street
Sacramento CA 95812

Submitted Electronically:
https://www.arb.ca.gov/lispub/comm/bccommlog.php?listname=capandtrade16

RE: Proposed 15-Day Modifications, Second Notice: Proposed Regulation and Mandatory Reporting Regulation

Dear Chairwoman Nichols and Members of the Board:

Agricultural Council of California (Ag Council) is writing in response to the April 13, 2017, Proposed Amendments to the California Cap on Greenhouse Gas Emissions and Market-Based Compliance Mechanisms Regulation (Proposed Regulation) and the Proposed Amendments to the Regulation for the Mandatory Reporting of Greenhouse Gas Emissions (Mandatory Reporting Regulation). A number of our member companies participate in the cap-and-trade program, and as a result, we appreciate the opportunity to comment on both the Proposed Regulation and the Mandatory Reporting Regulation.

Ag Council is a member-supported organization advocating for more than 15,000 farmers across California, ranging from small, farmer-owned businesses to the some of the world's best-known brands. Ag Council works tirelessly to keep its members productive and competitive, so that agriculture can continue to produce the highest quality food for the entire world.

Cap-and-Trade

Five years ago, the California Air Resources Board (ARB) launched the world's most complex cap-and-trade program. Although the program is working well in some areas, a number of challenges remain and a key test will come in the post-2020 period as the program cap continues to decline. As ARB looks to move beyond the initial iteration of cap-and-trade, it is vital for the program to carefully examine any policies that would drive away potential partners or sacrifice opportunities to decrease costs. The state's post-2020 policies should support the opportunity for new, emerging technologies and emissions control strategies that allow us to innovate. In light of Senate Bill 32's (Chapter 249, 2016) even more ambitious carbon reduction target, it is imperative that the regulation meet our environmental goals, while maintaining a strong economy and reducing leakage.

1121 L STREET, SUITE 304 SACRAMENTO, CA 95814
916.443.4887 PH 916.443.0601 FAX WWW.AGCOUNCIL.ORG

As ARB designs the next phase of cap-and-trade, Ag Council believes priority should be given to sustaining a reliable and stable supply of safe, high quality and affordable domestic food. California farmers and food processors produce food to the highest environmental and labor standards in the nation. However, complying with lower-emission and fuel-efficient standards comes at a financial cost. Higher costs put farmers and food processors at a disadvantage, since we are subject to global commodity markets and cannot simply raise prices to cover costs. Our main concern with ARB's cap-and-trade program is that it will lead to an increase in emissions leakage over time, as California-based entities continue to lose market share to competitors in other jurisdictions. A loss of market share could manifest in greater exposure to lower-cost imports and challenges in marketing California produced food. The declining emissions cap in the program allows the state to retain 100 percent assistance factors for regulated sectors and still achieve the greenhouse gas (GHG) emissions reduction targets, without increasing the risk of emissions leakage. We urge ARB to protect our food supply and create a flexible policy framework that will achieve cost-effective and technically feasible GHG emissions reductions.

Proposed Regulation

Ag Council would like to thank ARB for working with our staff and our members on a number of concerns outlined in previous comment letters related to definitions, benchmarks and setting new industry assistance factors.

Definitions and Benchmarks

We appreciate staff working with covered entities in the activities under Roasted Nuts and Peanut Butter Manufacturing to amend definitions in Section 95802, relating to pistachio and almond processing. These amendments help clarify covered activities that are performed under food manufacturing NAICS code 311911. We support the changes made here and to the related activities in Section 95891 and Table 9-1: Product-Based Emissions Efficiency Benchmarks.

We also found that ARB staff made appropriate changes to the benchmark activities and related definitions for NAICS code 31151 Dairy Product Manufacturing in Sections 95802 and 95891. In table 9-1, staff is proposing modified benchmarks for a variety of activities that better reflect engineering estimates and allocate emissions to each product more accurately. We appreciate staff reinstating existing benchmarks for allocation through vintage 2018 allowances and adding new products and benchmarks for vintage 2019 allowance allocation and beyond.

Post-2020 Industry Assistance Factor Calculations

Ag Council supports ARB's decision to delete table 8-3 and the text of section 95871(d). Delaying implementation of assistance factors for vintage year 2021 and beyond is important, since there are many outstanding issues with the previously proposed assistance factor methodology. We believe there was a real possibility that cap-and-trade would have increased emissions leakage under Attachment B in the first 15-day amended text released on December 21, 2016.[1] Data limitations and methodological choices used to conduct the leakage studies caused the authors of the studies to underestimate emissions leakage risk.

---

[1] https://www.arb.ca.gov/regact/2016/capandtrade16/attachb.pdf

Consequently, ARB proposed assistance factors that were too low to effectively mitigate emissions leakage for the food processing industry, as noted by Dr. Richard Sexton of U.C. Davis.[2]

We look forward to continuing to work with ARB to address our concerns and ensure that the program continues to function as intended.

Mandatory Reporting Regulation
In Section 95103 Greenhouse Gas Reporting Requirements, ARB is proposing to move up the deadline from September 1 to August 10 for when a reporting entity must complete and submit their third-party verification. Ag Council is concerned that shortening the timeframe for the assessments would impact our member's ability to work with the verifier to adequately address any questions and concerns. The reduced time period would also mean a rushed verification process, which could possibly lead to more errors. Our recommendation is to maintain the September 1 deadline.

Ag Council thanks ARB for considering our comments on the second 15-day amendments to the Proposed Regulation. We believe to achieve the 2030 goals for SB 32, the existing cap-and-trade program needs some improvement to minimize leakage in order for it to become a more meaningful program. Should you have any questions or concerns, please do not hesitate to contact Rachael O'Brien at (916) 443-4887 / Rachael@agcouncil.org.

Respectfully,

Emily Rooney

Emily Rooney
President

---

[2] Attachment A

Attachment A

Evaluation of the California Air Resources Board's Proposed Determination of Industry Assistance Factors for Post-2020 Compliance with AB32

Richard J. Sexton and Steven Sexton*

January 11, 2017

*Richard Sexton: Distinguished Professor of Agricultural and Resource Economics, University of California, Davis, One Shields Avenue, Davis, CA 95616, rich@primal.ucdavis.edu. Steven Sexton: Assistant Professor of Public Policy and Economics and Faculty Fellow of the Energy Initiative, Duke University, 201 Science Drive, Box 90245, Duke University, Durham, NC 27708, steven.sexton@duke.edu. The authors acknowledge funding for this report provided by the California League of Food Processors.

Contents

List of Figures

Executive Summary

In implementing a cap-and-trade program for greenhouse gas emissions pursuant to Assembly Bill (AB) 32, the California Air Resources Board (ARB) has freely allocated to firms substantial emissions allowances in initial compliance phases that continue until 2020. These allowance allocations are determined, in part, by output levels in previous periods, and, thus, serve to lower firms' marginal compliance costs. The generosity of initial allowance allocations is intended to aide regulated industries in transitioning to a regime of implicit emissions pricing. The allocations are also intended to satisfy the ARB's obligation under AB32 to mitigate emissions leakage "to the extent feasible." Leakage defines a phenomenon whereby production relocates outside of the regulated jurisdiction in order to avoid the costs of regulatory compliance. To the extent that regulated production is replaced by unregulated production, emissions are said to "leak" out of the emissions cap in the regulated jurisdiction. For globally mixing pollutants like greenhouse gases, these leaked emissions directly offset emissions reductions in the regulated jurisdiction, undermining the policy. In initial compliance phases, allowance allocations are also determined by ARB staff classification of industries according to their energy intensity and trade exposure (EITE).

As the ARB proposes to reduce allowance allocations for transition assistance, it directed staff to determine how allowance allocations might be changed to better mitigate emissions leakage. Staff subsequently commissioned three studies of industry-specific leakage risks intended to inform future industry assistance factors (AFs) that govern allowance allocations. A report by Meredith Fowlie, Mar Reguant, and Stephen Ryan, henceforth FRR, estimated for all EITE industries emissions leakage to international production (Fowlie et al. 2016b). Wayne Gray, Joshua Linn, and Richard Morgenstern authored a companion study (henceforth GLM) evaluating "domestic" leakage (Gray et al. 2016). A third report by Stephen F. Hamilton, Ethan Ligon, Aric Shafran, and Sofia Villas-Boas (hereafter Hamilton et al.) studied leakage from specific food-processing sectors (Hamilton et al. 2016). Subsequent to its receipt of these reports, ARB staff developed and revised modestly a proposal for determining AFs for compliance periods beginning in 2020.

This report reviews the methods and findings of FRR and GLM, as well as the use of these findings by ARB staff to develop new industry AFs. Though we are familiar with Hamilton et al.,

this report does not provide a review of that study because it is not employed by ARB staff in their proposal for determination of AFs. Each of the research teams commissioned by ARB staff to generate these reports is constituted by respected economists with relevant expertise. ARB staff are also well-qualified. Yet, we find their reports and the proposed determination of AFs to be deficient for the ARB's purposes due principally to the difficulty of the researchers' and staff's appointed tasks. But the deficiencies are also a consequence of limitations imposed by their methodological choices.

We have six main concerns about the reports and ARB's proposed determination of post-2020 AFs. They are:

1. The statistical models employed by FRR and GLM to estimate industry-specific leakage risks only allow leakage risks to vary across industries according to industry energy intensity. The approach does not allow leakage risks to vary according to the market structure of industries, even though market structure characteristics are known to affect leakage risks. By ignoring such characteristics as market power and price responsiveness of firm and market supply and demand, FRR and GLM are certain to estimate industry leakage risks with considerable error. Though other limitations of their analyses likely cause FRR and GLM to underestimate average emissions leakage risks, the exclusion of market characteristics from their statistical models assuredly yield overestimates of leakage risks for some industries and underestimates of leakage risks for other industries. Use of these estimates in the determination of industry AFs will induce capricious losses to many California firms, but may generate windfall gains to others.

2. Several limitations of the data employed by FRR and GLM in their statistical analyses likely cause them to underestimate the true risk of emissions leakage. The first of these limitations concerns the outcome variables they observe. Leakage rates are measured by policy-induced changes in the quantities of production in the regulated and unregulated jurisdictions. FRR and GLM, therefore, wish to observe how these quantities change in response to changes in input costs that mimic the input cost changes induced by California's cap-and-trade program. They can then infer how much emissions leakage the program is expected to cause. But FRR

and GLM do not observe quantities of production. Instead, they observe revenues (and values of imports and exports), which are defined by multiplying quantity times price. In response to input price changes, equilibrium prices and quantities move in opposite directions under almost any market conditions; revenues change less than proportionately with changes in quantities. Consequently, the FRR and GLM proxy for production changes is virtually certain to understate true changes in production quantities, which causes their estimates of emissions leakage rates to be too low.

FRR and GLM also use a poor proxy for the changes in firm costs induced by the cap-and-trade program. The program causes an indefinite increase in firm energy costs, and, consequently, in firm *expectations* of future energy costs in California. But FRR and GLM do not observe firm price expectations or permanent energy cost shocks that mimic the effect of cap-and-trade. Instead, they observe contemporaneous energy price changes that are likely to be idiosyncratic and transitory. Firms are unlikely to respond to such short-run fluctuations in input costs as they would respond to indefinite cost increases because production changes are costly. By relying upon contemporaneous energy prices as proxies for the changes in long-run energy price expectations that mirror the effects of the cap and trade program, FRR and GLM suffer substantial measurement error in their statistical model that also causes their estimates of emissions leakage to be too low.

The data employed by FRR and GLM in their statistical analysis are likely to induce two additional biases that cause them to underestimate emissions leakage. The first concerns the evolution of contemporaneous energy prices over their study periods. For much of the time period they study, contemporaneous energy prices declined because of the expansions of domestic energy extractions. Of course, the cap-and-trade program has the opposite effect on energy prices. Thus, by inferring firm responses to higher energy prices from observations on firm responses to lower energy prices, FRR and GLM assume firm responses to price increases and price decreases are symmetric. But firms can be expected to respond less to input cost reductions than to input cost increases because capacity constraints may preclude increased production in response to cost reductions whereas output reductions in response to cost increases are not similarly constrained. The final data limitation precludes FRR and GLM from isolating energy price changes faced only by California firms, like those due to cap-and-trade. Instead, they observe energy price changes faced by firms across the U.S. These changes in input costs are likely to induce changes in output prices that attenuate production responses. Changes in California energy costs, however, are unlikely to be accompanied by output price changes because California production is a small share of the national market in many industries. In the absence of output price changes, the changes in California firm production in response to California-specific energy cost changes will be greater than the changes of California production in response to national energy cost changes.

3. Although economic theory affords clear predictions that higher input costs in some jurisdictions will induce reallocations of production to jurisdictions with relatively lower input costs, all else being equal, theory does not clearly define the functional relationship between energy costs and production decisions. The econometrician, therefore, must exercise discretion in modeling the relationship. Consequently, FRR estimate 192 models and characterize elasticities derived from each of these models. However, they do not report the complete results of these models nor their implications for estimated emissions leakage. They and the ARB ignore model uncertainty demonstrated by these alternative models in subsequent analysis. GLM estimate only slight variations of their baseline statistical model and do not report any of the results. Moreover, neither FRR nor GLM report results from models that account for the dynamic adjustment of production to changes in input costs, even though theory would advocate for such models. FRR admit that leakage estimates are sensitive to modeling assumptions. In such instances, the public interest demands that model outputs be transparently conveyed. Unfortunately, the reader of FRR and GLM is largely left in the dark as to the consequences of modeling assumptions for AF determinations.

4. While modeling uncertainty confronting FRR and GLM is considerable, so, too, is the sampling uncertainty. It causes FRR and GLM to imprecisely estimate the parameters that define industry-specific leakage. FRR and GLM largely ignore the consequences of sampling uncertainty for their analysis, as does ARB staff in its subsequent use of the model

outputs. Astonishingly, GLM do not even report standard errors for their baseline parameter estimates. FRR depict for one of their 192 specifications the confidence intervals surrounding industry-specific estimates of production, import, and export responsiveness to energy price changes. These confidence intervals indicate that there is no statistically significant difference in emissions leakage risks across industries. Yet, in a manner that is arguably arbitrary, ARB staff proposes to use such parameters to differentially allocate allowances to industries.

5. FRR, GLM, and ARB staff repeatedly claim that their modeling assumptions generate estimates that should be interpreted as upper bounds on emissions leakage. Though we agree that some consequences of their methodologies may introduce upward bias in their estimates of emissions leakage, their claims to have identified upper bounds on emissions leakage are plainly not supported by economic or econometric theory. Various other modeling assumptions have countervailing effects that are likely to dominate any upward bias to yield emissions leakage estimates that are too low. Though FRR articulate some of the limitations of their analysis, GLM do not. And even FRR do not discuss any implications of modeling assumptions that would suggest they underestimate emissions leakage. The omission by FRR and GLM of any discussion about how methodological and data limitations may cause them to underestimate emissions leakage is curious and troubling.

6. In its proposed methodology for determining AFs, the ARB staff does exactly that which FRR caution against, saying it "amounts to pushing up against the limits of available data." FRR admit, "Given the noisiness of these estimates, we cannot estimate the transfer rate for any given industry with any degree of confidence." Yet, upon the basis of those estimates, ARB staff proposes to assign unique AFs to industries. They do so, however, only after adjusting specific model outputs from FRR and GLM that staff concluded were contrary to economic theory and staff members' own intuition. In particular, some model outputs suggest production and export responses to energy price increases are positive, though theory predicts they are negative. These results are plainly anomalous, a consequence of sample uncertainty and bias in their estimates. But in adjusting these parameter estimates, staff ineluctably admits a weakness of its approach: that all relevant parameter estimates are noisy, estimated

with error, likely to be biased, and, thus, also in need of adjustments, though less obviously so.

For these and other reasons enumerated in our report, we conclude:

- **Data limitations and methodological choices cause FRR and GLM to underestimate emissions leakage risk. Consequently, ARB proposed assistance factors are, on average, too low to effectively mitigate emissions leakage.**

- **The scientific evidence generated by FRR and GLM suggests that industry leakage risks are not statistically different at standard levels of scientific certainty. Therefore, it is arbitrary for ARB to assign high AFs to some industries and low AFs to others based upon these results.**

- **ARB staff manipulates some results of the FRR and GLM studies because they contradict accepted economic theory and staff judgement. These anomalous results are a consequence of the modeling approaches of FRR and GLM and of the considerable uncertainty in estimates of emissions leakage. In manipulating these results, the ARB admits that all of its estimates of industry-specific leakage are subject to similar errors.**

Additional panel-econometric work of the type performed by FRR and GLM could yield improved estimates, e.g., through modeling dynamic production response. Revised reports by FRR and GLM could also provide greater transparency in order to permit further inquiry as to the implications of modeling assumptions for estimation of leakage risk. In several respects, the reports do not meet the professional standard for transparency, even though the public interest in peer review and replication is considerable given the immediate and direct impact these studies are likely to have on an important public policy.

Still, the panel-econometric approach to estimating industry-specific leakages is likely inadequate because of the vast heterogeneity in production technologies and market structures across regulated California firms and industries and because data to account for these differences are apparently lacking. Thus, we conclude that industry-specific analyses could generate more credible estimates of leakage risks by accounting for differences in market characteristics across industries. Analysis such as that undertaken by Hamilton et al. can account for industries' heterogeneous market structures and dynamic adjustments to persistently elevated energy costs. The credibility of such analyses will depend upon the availability of reliable estimates of market structure parameters. We also note that FRR, themselves, undertook a recent analysis of leakage risk for the U.S. Portland cement industry (Fowlie et al. 2016a). In their study of the cement industry, they highlight their reliance upon a dynamic, structural model for estimating production responses. Such a model contrasts markedly with the reduced-form, static approach that characterizes the baseline models in their and GLM's reports to the ARB.

1 Introduction

California's "Global Warming Solutions Act" (Assembly Bill 32) and subsequent legislation establish statewide targets for greenhouse gas emissions reductions. The laws require that greenhouse gas (GHG) emissions not exceed 1990 levels by 2020 and that they decline a further 40% by 2030. Charged with implementing regulations to achieve state emissions targets, the California Air Resources Board (ARB) introduced in 2012 a cap-and-trade program that limits allowable emissions from approximately 450 entities in regulated sectors collectively accountable for 85% of total emissions. Via emissions allowance auctions, the program establishes a price of emissions, i.e., a carbon price, that consequently raises the cost of production by regulated firms. To the extent carbon emissions are unpriced or underpriced outside of California, then the program may induce relocation of emitting production to other states or countries where it is cheaper, *ceteris paribus.* Because GHGs are globally mixing pollutants, the emissions from increased production outside of the state offset emissions reductions within the state, a phenomenon known as leakage.[1]

In recognition of the potential for emissions leakage, AB 32 compels the ARB to minimize leakage "to the extent feasible" in its operation of the cap-and-trade program and other related rule making. Consequently, and in order to facilitate the transition to emissions pricing, the ARB has freely distributed emissions allowances to firms according to their production levels and leakage risks. The ARB presently infers leakage risk from calculations of industry-specific emissions intensity and trade exposure.[2] The free allowances effectively lower the marginal cost of emissions, and, thereby production, for those firms receiving allowances. They are expected to reduce leakage by lessening the policy-induced cost advantage of outside production relative to production within the state.

In contemplating leakage mitigation efforts beyond 2020, the ARB directed staff to "investigate potential improvements" to its system of allowance allocation in order to better minimize leakage.[3] Staff subsequently commissioned three studies to inform the development of industry-specific "assistance factors" (AFs) that will govern allowance allocations in a future

[1]Emissions are said to "leak" out of the cap on emissions in the regulated jurisdiction.
[2]See https://www.arb.ca.gov/regact/2010/capandtrade10/capv4appj.pdf.
[3]See https://www.arb.ca.gov/cc/capandtrade/meetings/20161021/ct-af-proposal-102116.pdf.

with less transition assistance and a stricter emissions cap. One of these studies, conducted by Meredith Fowlie, Mar Reguant, and Stephen Ryan (hereafter FRR), focused on international leakages; another by Wayne Gray, Joshua Linn, and Richard Morgenstern (hereafter GLM), focused on leakages to other states (i.e., domestic leakages); and a third by Stephen F. Hamilton, Ethan Ligon, Aric Shafran, and Sofia Villas-Boas (hereafter Hamilton et al.) studied leakage from specific food-processing sectors (Fowlie et al. 2016b, Gray et al. 2016, Hamilton et al. 2016).

In October 2016, subsequent to the completion of these studies, ARB staff prepared an informal proposal to employ empirical model outputs from the FRR and GLM studies in developing AFs for regulated industries. The staff proposal and subsequent updates issued in November 2016 largely ignore the results of Hamilton et al., while undertaking various adjustments to specific model outputs from FRR and GLM that ARB staff deem counterintuitive or anomalous.

The purpose of this report is to review the FRR and GLM studies, as well as the staff proposal to employ them in implementing industry AFs for post-2020 allowance allocation. Authors of FRR and GLM are respected economists with relevant expertise, and the ARB staff are well-qualified; however, we conclude that limitations of the panel-econometric approach they employ result in deficiencies that render them unsuitable for determining industry AFs. Appropriate allocation of emissions allowances is critical to the ARB mandate to minimize leakage and, indeed, to the success of California's climate-change program. The determination of leakage potential across industries and of mitigation measures is essential for the effective implementation of regional climate-change policies.

Regulated California firms face energy costs that are increased indefinitely due to the pricing of GHG emissions. These persistently higher input costs eventually reduce the production of a regulated firm under any standard model of production and competition. But the timing and magnitude of production decreases depend upon the interplay of a complex set of factors involving production technology, fixity of certain capital inputs, demand for firm products, contractual relationships with input suppliers and downstream buyers, and the degree of competition from domestic and international rivals. These considerations interact through the marketplace to determine output and emissions reductions for California firms and the extent to which both are supplanted by output and emissions from firms outside California. These many considerations imply that production responses and leakage due to regional climate policy will vary across industry, as FRR, GLM, and ARB staff concede by virtue of their efforts to determine industry-specific AFs. Accounting properly and fully for the industry-specific factors governing production responses, and, subsequently, leakage, constitutes a herculean undertaking.

In order to demonstrate the difficulty of the task before ARB staff and to highlight the limitations for staff purposes of the FRR and GLM empirical analyses, it is instructive to consider the data one would ideally bring to bear in estimating industry-specific leakage risks. First, one would want to observe an exogenous increase in the long-run energy price expectations of a subset of California firm and plant managers within each industry. Second, one would want to observe output from these firms over a period of more than a decade and relate their output to that of other firms in California to determine how energy price expectations affected in-state production in the long run. Third, one would ideally observe production decisions of firms elsewhere around the world in order to determine the extent to which diminished California output is replaced by production elsewhere. Finally, because the magnitude of emissions leakage depends upon the emissions intensity of reallocated production, one ideally observes the emissions intensity of all firms whose production responds to the exogenous change in energy price expectations among the select California firms.

Armed with these data, one could compare the short- and long-run production decisions of California firms perceiving higher energy prices to those of firms with unchanged energy price expectations in order to assess the causal effect of higher energy prices (and price expectations) on California production for each industry. Moreover, one could determine how production outside California responds to changes in energy price expectations for select California firms in order to determine what fraction of production changes within the state are offset by production changes elsewhere.

These ideal data are hard to come by. For instance, an exogenous (or "as good as random") change in energy price expectations among some firms but not others is impossible to imagine outside of a randomized control trial in which some firms within each industry are "treated" with permanently higher energy costs while other firms face unchanged energy prices. In other words, these data would need to be created in an experimental setting that likely would never be politically viable even if researchers were sufficiently ambitious. Unsurprisingly, FRR and GLM do not undertake such an experiment. They also do not observe output from any plants, firms, or industries, let alone output from all worldwide production facilities. Though these data surely exist, at least within the firms, they are privately held and often closely guarded. FRR and GLM also do not observe any plant emissions intensities, though these exist for some plants in the U.S. They do not exist for other plants in the U.S., and are even scarcer for facilities elsewhere around the world. In the absence of these data, FRR and GLM necessarily make assumptions in order to make inferences about the production impacts of California climate policy, and their analyses are consequently limited by the appropriateness of these assumptions. Many of these assumptions have important affects on their empirical results, and, yet, are inconsistent with economic theory.

In completing this review, we are cognizant that the ARB has not commissioned its own review of FRR and GLM, or of its intended use of those studies in setting industry AFs.[4] To our knowledge, neither has the ARB staff proposal for future allowance allocations been vetted by the ARB's Economic and Allocation Advisory Committee. Our own review is constrained by the information that is contained in the FRR and GLM reports published on the ARB website and by staff's published draft proposal. The data and programs employed by FRR and GLM in conducting their analyses are not publicly available, precluding efforts to replicate their findings or to interrogate their data, as is essential to scientific inquiry. In recognition of the importance of replication, the American Economic Association has adopted a policy of publishing papers "only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication."[5] No lesser standard should apply to research that is expected to directly and immediately influence policy that will impact production and employment in California and affect billions of dollars of transfers to and from California firms.

Though this inquiry is limited to only the analysis contained in the FRR and GLM reports, and not the underlying data, we nevertheless undertook a concerted review of their empirical

[4]We are aware of another independent review by Bernstein et al. (2016), of NERA Consultants, as well as comments to the ARB on the studies by interested parties available here: https://www.arb.ca.gov/lispub/comm2/bccommlog.php?listname=ctleakagestudies-ws.

[5]https://www.aeaweb.org/journals/policies/data-availability-policy.

methods and of the comportment of those methods with economic theory. We reach several conclusions. First, while the FRR and GLM studies constitute valuable — and even valiant — efforts to understand how regional cap and trade programs impact the regional economy and global greenhouse gas emissions, they are not appropriate for setting industry-specific allowance allocations. On the basis of their own report, FRR appear to agree.

Second, we dispute the repeated contention of each study and the ARB staff proposal that their estimates of emissions leakage constitute "upper bounds" on likely leakage and that they make "conservative" assumptions that favor industries and overly guard against leakage. These claims of conservatism convey a belief, misguided in our view, that to the extent the cap-and-trade program AFs are erroneous, they err on the side of minimizing harm to California industries. To the contrary, these studies likely understate the risk of leakage, and ex post adjustments made by ARB staff do little to mitigate against it.

Third, more generally, we do not believe that the various modifications, adjustments, and re-analysis of FRR and GLM outputs performed by ARB staff do anything to overcome the deficiencies of these studies or to improve their appropriateness in determining AFs. Instead, they admit the fundamental problems with the FRR and GLM studies and the ARB staff proposal for setting future allowance allocations.

This report proceeds with an enumeration of our chief critiques of the FRR and GLM studies. We begin by enumerating our theoretical and conceptual critiques of their empirical approaches before turning to a discussion of their implementation and reporting of results. We then evaluate the ARB staff proposal for determining AFs in the post-2020 compliance period. Before concluding with limited recommendations for the ARB, we challenge the notion that the FRR and GLM studies and ARB staff proposal yield a conservative, or industry friendly, approach to leakage mitigation. We do so by enumerating assumptions and empirical limitations that likely cause the estimates of leakage to be too low and the proposed AFs to be insufficient to effectively mitigate leakage as the ARB is compelled to do.

2 Critiques of theoretical foundations

2.1 Firms are expected to respond differently to contemporaneous price changes than to changes in long-run price expectations like those induced by carbon pricing.

Basic production analysis suggests that firms can moderate, over time (i.e., in what economists refer to as the long run), the negative profit impacts of an input price increase by substituting other inputs for those that become costlier. However, such substitutions, or adjustments, are themselves costly because of (a) transactions costs associated with disrupting the input mix and production schedule, (b) ex ante commitments to buy inputs, (c) ex ante contractual commitments to sell outputs, (d) presence of inventories of inputs, and (e) the sunk nature of capital investments in the short run which limits firms' abilities to adjust production by altering capacity, relocating production across plants, etc. Firms are unlikely to incur these adjustment costs in response to input price changes that are perceived to be idiosyncratic or transitory. Changes in capital, production, and input mix are likely to occur, however, in response to expectations of persistent changes in input prices, like those induced by climate policy in California.[6] Davis and Kilian (2011) make the same observation in the related context of gasoline taxes:

> Price changes induced by tax changes are more persistent than other price changes and thus may induce larger behavioral changes.

Hence, any econometric assessment of production impacts of climate policy should relate production decisions to changes in long-run energy price expectations.[7] FRR admit as much, noting, "We therefore need to use variation in relative operating costs that mimics the effect of [cap-and-trade] compliance costs to estimate the effects we are interested in."

Yet neither FRR nor GLM observe exogenous changes in energy price expectations, or any changes in energy price expectations at all. Instead, they observe some contemporaneous energy prices and estimate others. To justify their reliance upon contemporaneous energy prices rather

[6]Pindyck and Rotemberg (1983) develop the dynamic production theory pertaining to an energy price shock in the presence of rational price expectations and costs of adjusting the input mix.

[7]Bernstein et al. (2016) make a similar point in their comment on the ARB-commissioned studies.

than price expectations, the authors must rely implicitly on one of two assumptions. Either contemporaneous prices must be assumed to be an adequate proxy for price expectations, i.e., that the two time series are sufficiently correlated, or contemporaneous, and potentially transitory, price changes must be sufficient to induce the production changes that theory predicts attend changes in long-run price expectations, like those induced by climate policy. The former assumption is empirically dubious, as we argue below, while the latter conflicts with accepted theory, e.g. Pindyck and Rotemberg (1983) and Davis and Kilian (2011).

Though GLM do not explicitly state the assumption that justifies their reliance upon contemporaneous prices, FRR admit (p. 28):

> Ultimately, our ability to extract policy implications from this analysis is predicated on the assumption that a careful analysis of how firms have historically responded to persistent changes in relative energy costs can inform our understanding of how a carbon price would impact industrial production and international trade flows.

FRR implicitly concede that price expectations and not contemporaneous prices govern the relevant firm responses. They do so by appealing repeatedly to their observation of production outcomes in response to "persistent," "significant," and "sustained" price changes, and, particularly, a secular decline in domestic energy prices due to the shale energy boom that began in earnest around 2008. FRR argue that "some portion of this variation can be used to isolate and estimate how changes in relative industrial operating costs (i.e., domestic versus foreign) have impacted domestic production and associated trade patterns across different industries." Though unstated, the implication is that long-run price expectations during this period matched the decline in prices because of the persistence of the contemporaneous price declines. If true, FRR and GLM could evaluate how production outcomes responded to those observed prices in order to infer how firms respond to carbon pricing if they further assumed that firm response to input price increases and price decreases is symmetric. This additional assumption is, itself, dubious as we argue later in this report.

While the largely overlapping time periods FRR and GLM study, 1993-2012 and 1989-2009, respectively, were, indeed, characterized by dramatic domestic energy price variation, the period is ill suited to the purposes of FRR and GLM for at least several reasons. First, the period of persistent energy price declines that FRR attribute to "sustained growth in domestic extraction" is contaminated by the Great Recession that induced the most dramatic price variation of the period. Natural gas and oil prices had been climbing rapidly and steadily in early 2008 before abruptly changing course in the latter half of the year at the onset of the recession. U.S. GDP would not recover to pre-recession levels until mid-2011. The evolution of these prices and of U.S. GDP are depicted in figure 1. As we argue later in this report, in so far as FRR and GLM would invoke the secular decline in energy prices to validate their empirical approach, then they would be compelled to concede that the Great Recession and the subsequent slow economic recovery confounds their assessment of production responses to energy price declines.

Intuitively, demand for energy declined due to the Great Recession during the same time period as the increase in energy supply due to the fracking boom. Muted production responses to reduced energy costs may be due to low price sensitivity of production to energy costs or due to contemporaneous reductions in demand. Both FRR and GLM employ statistical techniques to mitigate the potential for the Great Recession to confound their analyses, i.e., they employ year fixed effects to "detrend" the data. But the same statistical techniques that remove GDP trends also remove energy price trends. Moreover, the recession impacted industries to varying degrees, with less energy-intensive service sectors emerging less affected than manufacturing. These sector or industry-specific effects of the recession are accounted for to only limited extents in FRR. We revisit both these points in Section 3 of this report.

Second, the most dramatic energy price changes attributable to shale energy production in the U.S were only beginning to be realized at the end of the FRR study period and well beyond the end of the GLM study period. As figure 2(a) reveals, natural gas prices fell to $1.83 per million BTUs (MMBtu) in 2009 (the end of the GLM study period) at the onset of the Great Recession, a drop of nearly $5 from a year earlier. These changes could have been viewed as transitory, reflecting the drop in aggregate demand that would eventually rebound. The decline in gas prices attributable to the shale production shock and reflecting a new era of domestic hydrocarbon recovery arrived in 2012, at the end of the FRR study period. Thus, FRR and GLM are unlikely to have observed any production changes resulting from sustained lower natural gas prices let alone lowered price expectations due to the shale boom. GLM largely missed the secular price decline entirely. Likewise, and as depicted in figure 2(b), oil prices largely remained high, above $80 per barrel, until 2014, well beyond the FRR and GLM study periods. Oil prices plummeted only after the Organization of Petroleum Exporting Countries (OPEC) declined to curtail output in response to dramatic expansion of U.S. production.

Third, the period of sustained low energy prices observed by FRR and GLM was a period of tremendous price uncertainty (EIA 2010, EIA 2012b). Though energy prices reached historic lows during the study period — natural gas spot prices reached a low of $1.82/ MMBtu in 2012 — and remained low well beyond the study period, these declines were largely unanticipated.[8] As recently as 2012, the EIA anticipated gas prices would barely dip below $4 before increasing steadily throughout the projection period at 2.3% per year (EIA 2012a). In 2007, the EIA expected gas prices to average $5 in 2013 and to increase modestly to about $6 by 2030. Never during the study period, did the EIA reference case project prices as low as those that were realized since 2012. The EIA also projected rising world oil prices. In fact, the EIA reference case projected West Texas Intermediate (WTI) crude prices of $117 per barrel in 2015 and $127 in 2020 (EIA 2012a). The WTI closed 2015 at $36.

Energy futures contracts also indicate that energy prices were not expected to decline substantially and persistently during the study period. Figures 3 and 4 depict futures prices for natural gas and oil, respectively. Each figure shows the evolution of prices for generic 60-month, 48-month, 36-month, and front-month futures. Assuming efficient markets for these futures, they reflect at any instant the market expectation of energy prices at delivery in the future. Thus, front month futures reflect expected prices 30-days hence, effectively reflecting contemporaneous prices. Sixty-month futures reflect expectations for prices five years in the future. At the onset of the secular price declines in 2008, 5-year futures traded at above $8/ MMBtu, reflecting market expectations of a high gas price future throughout the remainder of the FRR study period. In 2009, four-year futures traded around $7, exhibiting little anticipation of the impending slump. And in 2010, three-year futures anticipated prices around $6 throughout the FRR study period. Similarly,

[8]Henry Hub natural gas sport prices were obtained from the EIA at https://www.eia.gov/dnav/ng/hist/rngwhhdd.htm.

oil prices were expected to remain above $80 per barrel well beyond the FRR study period, as demonstrated by the 3, 4, and 5-year futures prices in the years preceding 2012.

The decline in energy prices that persisted through 2016 was largely missed by market participants because of (1) an unanticipated sluggish recovery from the Great Recession that caused demand to rebound slowly; (2) unexpected improvements in shale extraction technology; and (3) decisions by domestic and foreign producers to continue pumping oil and gas amid low energy prices (Furman 2016). Joskow (2013) characterized the then-current and future growth in production of shale gas as "dramatic and largely unanticipated," and he credited it with "dramatically chang(ing) the energy future of the United States and potentially of the world."

It is notable that even in its 2012 energy projections, the EIA assumed in its reference projections historic rates of productivity gains in shale energy extraction. But falling prices imposed discipline on producers who dramatically improved efficiency, reducing the rig time to drill wells and increasing the production per well by expanding the lengths of laterals, drilling multiple wells per pad, and strategically fracking adjacent wells to maximize production. These innovations allowed production to continue at prices below which industry analysts expected it to be unprofitable. Firms also directed attention to more productive wells, and they drilled wells that were not "in the money" in order to maintain leases (Denning 2010, Ailworth 2015, Friedman 2014, Gold 2015, Olson and Ailworth 2015). The EIA (2012b) characterized the impact of uncertainty in shale gas production thusly:

> Considerable uncertainty exists regarding the size of the economically recoverable U.S. shale gas resource base and the cost of producing those resources. Across four shale gas resource scenarios from the Annual Energy Outlook 2012, natural gas prices vary by about $4 per million British thermal units in 2035, demonstrating the significant impact that shale gas resource uncertainty has in determining future natural gas prices. This uncertainty exists primarily because shale gas wells exhibit a wide variation in their initial production rate, rate of decline, and estimated ultimate recovery per well.

Low oil prices persisted longer than many industry observers expected because restraint among OPEC producers was surprisingly absent. Furman (2016) called the dramatic decline in oil prices "the biggest surprise in the last year in the global economy."

Low domestic energy prices also reflected constrained energy commodity movements within and across borders that were partly induced by policy decisions, and, thus, subject to revision. Importantly, while low domestic natural gas prices provided ample opportunity to arbitrage international prices, infrastructure and regulation constrained product movement to more profitable international markets, thereby preserving the low domestic prices. The U.S., long an importer of liquefied natural gas (LNG), was not equipped with liquefaction terminals to ship natural gas overseas. Development of such terminals, costing billions of dollars each, was also constrained by an uncertain and slow federal approval process that made the persistence of the domestic natural gas price advantage uncertain (Economist 2013). The IEA (2009), for instance, projected U.S. gas prices to rise 7-19% by 2025 due to gas exports.

Thus, the claim that long-run price expectations matched persistent commodity price declines is a tenuous one at best. To the extent contemporaneous prices do not reflect price expectations, then their use in estimating emissions leakage likely induces classical measurement error that is known to attenuate estimates of marginal or causal effects (e.g., Griliches and Hausman 1986). That is, the FRR and GLM estimates of leakage are biased toward zero because they attribute changes in production outcomes to dramatic price changes that do not bear on firm responses and may bear little resemblance to movements of price expectations that do factor into production decisions, at least in the long run. Figures 3 and 4 illustrate this point. Front-month futures, which closely reflect contemporaneous prices, exhibit far greater price volatility than the distant futures contracts that reflect the markets expectations for long-run prices.

2.2 Use of values of production, exports, and imports understate the reductions in output caused by higher input prices.

Generalizing FRR (see p. 38), emissions leakage is defined as the product of market transfer rates and emissions intensities. Specifically, emissions leakage to any jurisdiction outside California is defined as the emissions intensity of industry production in that jurisdiction times the rate at which a reduction in California production translates into an increase in production in that jurisdiction.

Thus, to estimate industry-specific leakage, one ideally observes changes in industry production within California and within other states or countries. Yet neither FRR nor GLM directly observe output responses of any industry in any jurisdiction. Instead, FRR and GLM observe changes in sales revenue — the product of price times quantity sold — and rely upon these changes to proxy for changes in output. However, by conflating output responses and output price changes, these proxies for production changes are likely to understate the true magnitudes of output changes, and, consequently, to understate emissions leakage.[9]

The error introduced by relying upon value-of-production measures of firm response instead of directly observing output may be substantial. At the industry level, demand slopes downward as a function of price, regardless of whether the industry is perfectly competitive, as FRR and GLM assume, or evinces some form of imperfect competition. This means that a market supply shock, for example due to a change in energy price expectations, impacts both prices and quantities in opposite and offsetting directions.

Because FRR and GLM rely on energy price changes driven largely by national and global trends, those price changes will have been experienced to some degree by all firms operating in the industry and are virtually certain to affect output prices through the normal workings of the marketplace. Quite simply, higher energy costs shift the industry supply relationship, causing a price impact that is opposite the production impact. Consequently, sales or value-added variables employed by FRR and GLM assuredly understate the response of output and the potential for leakage.

A related bias is introduced by reliance upon national (and international) energy price variation to estimate production responses from price changes affecting only California firms, like the state's cap-and-trade program. California generally has a small share of the national market for energy-intensive industries (Bernstein et al. 2016).[10] This means that higher California energy prices induced by climate policy are unlikely to generate much, if any, impact on market prices. The

[9]FRR also observe changes in the values of imports and exports and GLM also observe changes in value added and in quantities of labor inputs.

[10]According to Bernstein et al. (2016) (p. 25): Output of California industries in the study range from 0.6% to 15.5% of the total national output of the industry, according to IMPLAN data. Their footnote 25 contains estimates of CA share of national market for energy-intensive industries.

impact on California firms will be experienced almost exclusively through outputs because output prices will remain largely unchanged. Hence, even if FRR and GLM were correctly observing outputs, the leakage risk for a California-specific energy price change will be greater — perhaps considerably greater — than what FRR and GLM. The industry-level output responses they observe are reduced due to the offsetting change in output prices.

Figures 5(a) and 5(b) illustrate the point. Figure 5(a) illustrates a competitive industry, and figure 5(b) illustrates a California firm operating in that industry. An initial equilibrium is depicted with market price, p_0, and quantity, Q_0, and firm output, q_0. A permanent national increase in energy costs increases marginal costs for all firms in the industry similar to what is shown in FRR's figure 2. Marginal costs for the California firm shift from MC_0 to MC_1 and industry supply (the sum of marginal costs across firms) shifts by a commensurate vertical distance, from S_0 to S_1. The change in sales revenue at the industry level, as would be measured by FRR and GLM is $P_1Q_1 - P_0Q_0$, which may be positive or negative, and in any event understates the industry's reduction in output and emissions. The California firm's reduction in output is from q_0 to q_1; it is muted by the higher price achieved through the market. If FRR and GLM were able to measure the industry output change, Q_1-Q_0, correctly, it would be a good proxy on a percentage or elasticity basis for the change experienced by the California firm, $q_1 - q_0$. However, $P_1Q_1 - P_0Q_0$ is biased downward as a proxy for $q_1 - q_0$. Moreover, California's cap-and-trade program only raises energy prices for California firms, so the industry supply shift will be negligible; California has a small share of the total market. This means price effectively remains at P_0, and the California firm's output reduction is from q_0 to q_2. Thus, even if FRR and GLM observed Q_1-Q_0 instead of $P_1Q_1 - P_0Q_0$, their estimate of California firms' output response to a California-only energy price change would be too low.

2.3 Firm responses to input price changes occur over time as inputs become substitutable and capital becomes mobile.

Just as theory holds that firms respond to changes in long-run input price expectations rather than transitory price shocks, so, too, does theory predict that firms respond to changes in expectations gradually, over time. Instantaneous production responses may be muted by ex ante commitments to buy inputs or to sell outputs, or by inventories, or by the sunk nature of capital investments in the short run that limit firm abilities to alter capacity and relocate production across plants.[11]

FRR agree (p. 25):

> Industry-level response to production cost shocks can take time to play out completely; many industries have large capital shares that adjust slowly. Documenting industry response over short time scales may capture only a fraction of the response to changes in relative operating costs (e.g., short-run re-optimization over inputs to production).

Likewise, in other work these same authors note that over longer time frames, "firms can alter their choice of production scale, technology, entry, exit, or investment behavior in response to an environmental policy intervention" (Fowlie et al. 2016a).[12] Indeed, FRR find in their empirical modeling that industries characterized by low capital intensity, which they deem more mobile, exhibit greater short-run production responses to energy price changes than do capital-intensive industries. By definition, all capital is mobile in the long run. Therefore, and contrary to the findings of GLM, long-run production responses to a permanent energy price change can be expected to exceed short-run responses.[13]

Because firms respond over time to changes in input prices, static econometric models, such as the baseline models in FRR and GLM, are expected to understate firm responses to carbon pricing by ignoring the dynamic adjustment process. FRR estimate 192 distinct specifications of the equation relating production outcomes to energy prices, admitting uncertainty over which model best defines the relationship. They indicate that lagged production outcomes that permit dynamic firm responses to contemporaneous energy prices are included as independent variables, or covariates, in some of these specifications. But nowhere are the parameter estimates of these

[11]As we discuss in detail among our empirical critiques in the subsequent section, the authors' use of time fixed effects absorbs secular energy price trends, leaving them with only idiosyncratic price shocks from which to identify firms' responses to climate policy.

[12]Fowlie et al. (2016a) also note, "The welfare effects of a market-based emissions policy can look quite different across otherwise similar static and dynamic modeling frameworks."

[13]The GLM assertion is apparently motivated by a belief that firms will invest in greater input-use efficiency in the long run, thereby reducing sensitivity to energy prices. This substitution of capital for energy inputs is precisely the kind of input substitution theory predicts over sufficient time horizons. It is far from obvious, however, that such substitutions generate on net less production response in the long run than in the short run.

models identified in FRR. Consequently, the reader is unable to evaluate how well these dynamic models fit the data or what they imply about emissions leakage.

GLM do not even estimate a dynamic model. Instead, GLM estimate a "long differences" regression (see equation 3) of the sort employed in empirical macroeconomics to estimate longer-run impacts of independent variables, while purging estimates of some omitted variables bias. These long-difference estimates subtract 5-year lagged outcomes from contemporaneous outcomes and regress the difference in outcomes on the difference in contemporaneous and five-year lagged energy prices. The estimated production responses from these long-difference (long-run) models in GLM are smaller — generally much smaller — than the production responses estimated in the short-run, fixed effects models that employ the "within" transformation to purge omitted variables bias.[14] GLM infer from this result that production outcomes are less responsive to energy price changes over five years than they are to contemporaneous price changes, i.e. short-run production is *more* sensitive to input price than long-run production. This result is surprising, and contrary to accepted theory as we, and, seemingly, FRR, understand it.

GLM contend that this result reflects greater opportunities for improvements to input-use efficiency overtime as old capital is replaced, e.g., with energy efficient capital.[15] Given the discordance of this result with accepted theory, however, it is more likely that attenuation bias (or measurement error bias) causes the long difference estimates of production responses to be smaller than the estimates from the fixed-effects model. Measurement error is known to introduce considerable inconsistency in within estimators and difference estimators that exacerbate "attenuation bias" or "errors in variables bias" in order to eliminate omitted variables bias (Griliches and Hausman 1986).

Because GLM and FRR observe only contemporaneous energy prices, and not energy price expectations, their estimates are all likely to suffer from classical measurement error that

[14]See Wooldridge (2010), p. 265, or Angrist and Pischke (2008), p. 165. In order to remove time invariant, individual-specific factors and common secular trends that may bias estimates of causal effects, individual means and time means of dependent and independent variables are subtracted from each observation so that estimation proceeds on deviations from means.

[15]Although this is a hoped-for response to higher energy prices due to the cap-and-trade program, we should remember that, as FRR note pointedly, the data they and GLM analyzed do not pertain to a time period of steadily rising energy prices, which would have motivated firms to adopt energy-efficient technologies.

biases estimated causal, or marginal, effects towards zero. Both difference estimators and within estimators commonly employed to remove omitted variables bias worsen the attenuation bias caused by measurement error because they remove variation in the proxy variable (here energy prices) due to variation in the unobserved variable of interest, leaving primarily variation due to noise in the correspondence between the proxy variable and the variable of interest. Griliches and Hausman (1986) establish that this attenuation bias is greater in first difference estimators than in fixed effects (or within) estimators.

Because long differences are likely less serially correlated than first differences, the attenuation bias is likely lower in the long difference estimator than in the first difference estimator, yet it may nevertheless remain greater than the bias in the within estimator. Consequently, the long difference estimator provides unreliable inference as to the magnitude of production responses to energy prices, and any effort to compare the magnitudes of short-run responses derived from the within estimator to the long-run responses derived from the difference estimator are erroneous because of likely differences in attenuation bias. Thus, contrary to the claim of GLM, one cannot conclude from the relative magnitudes of estimated production responses in their short and long-run models that production responses are smaller in the long run than the short run. And, therefore, their reliance upon their short-run parameter estimates in determining domestic emissions leakage does not reflect conservatism as they contend. More generally, measurement error also surely plagues the estimation in FRR, as does their reliance upon the within transformation amid measurement error. All model estimates in FRR and GLM are likely biased toward zero, and, therefore yield leakage estimates that are too small, rather than too large.

2.4 Market structure and heterogeneous production processes likely cause considerable heterogeneity in input-price sensitivity across industries that is not captured by an assumption of perfect competition or homogeneous response.

The empirical analyses of FRR and GLM are intended to inform the ARB's determination of industry-specific AFs. Yet FRR and GLM do not estimate industry-specific responses to the energy price changes. Rather, FRR admit only heterogeneity in energy intensity, allowing a common energy price response to deviate only according to industry energy intensity. GLM allow price responses to vary according to five ad hoc groups of industries and according to energy intensity. They ignore other forms of heterogeneity that may cause industry-specific price responses, and, thereby, emissions leakage, to vary markedly across industries. They ignore differences in competitive conditions, product differentiation, supply chains, and other firm characteristics that are likely to generate heterogeneous carbon price impacts within and across industries.

Bushnell and Humber (2015), in a paper cited by FRR, provide an illustration. They find that the nitrogenous fertilizer industry, an energy-intensive and trade-exposed industry eligible for leakage protection, exhibited minimal response to large changes in natural gas prices even though natural gas is the key input in the production process. They attribute this outcome to market power and a domestic cost advantage relative to international competitors. Bushnell and Humber conclude that leakage protections to the nitrogenous fertilizer industry would be "nearly completely unnecessary in terms of its stated goal of protecting local producers, while at the same time constituting a substantial windfall to those same producers."

It is also likely that two firms facing identical energy price changes and characterized by identical energy intensities could have drastically different sales responses due to differences in the structure of demand and competition they face. Figures 6(a) and 6(b) illustrate this point using two standard models from economic theory. Figure 6(a) depicts a prototype monopolist who produces where marginal cost equals marginal revenue, as given by output Q_0 in 6(a) prior to imposition of a cap-and-trade program. Cap-and-trade increases the firm's energy input costs, shifting up its marginal cost, reducing output to Q_1 and increasing price from P_0 to P_1. The cap-and-trade program reduces output of the firm by internalizing the negative externality from GHG emissions. No direct leakage occurs given this firm's monopoly position, and no AF should be provided.

Figure 6(b), in contrast, presents a case of price-setting duopolists (two competing firms that jointly comprise the selling side of a market) who sell a homogeneous product. The equilibrium in this setting is for firms to set price to the level of the their marginal costs if those costs are identical, or, for the firm with a cost advantage relative to the rival to set price just below the rival's marginal cost, thereby driving the high-cost rival from the market. Figure 6(b) shows the firms facing identical marginal costs, c_0, prior to the imposition of the carbon price on the California firm. They share the market output, each producing $\frac{1}{2}Q_0$, with market price $P_0 = c_0$. When the California firm faces a carbon price, its marginal costs increase to c_1, where the vertical shift in marginal costs in figure 6(b) is set to equal that of the monopolist in figure 6(a). The (Bertrand-Nash) equilibrium price is now $P_1 = c_1 - \epsilon$, where ϵ is a small positive number, i.e., the low-cost rival to the California firm exploits its cost advantage to undercut any price the California firm is able to charge without incurring a loss. Here the California firm incurs a drastic reduction in output due to the carbon price, and nearly all of it is transferred as leakage to the unregulated firm.

Because the two firms illustrated in this figure have identical energy intensities, they would likely receive identical AFs under the ARB staff proposal. Clearly, however, their production responses to the carbon price are drastically different. Admittedly, neither the pure monopoly model nor Bertrand duopoly model is necessarily an accurate depiction of any of the industries studied by FRR or GLM, but neither is the competitive model upon which they exclusively rely.

As FRR note (p. 25), "Estimates of average effects can mask economically significant heterogeneity in industry and firm-level responses and impacts." This is a severe limitation of studies intended to inform the development of industry-specific AFs. Alone, the failure to account for the many dimensions that will cause heterogeneous responses among regulated California plants is almost certain to result in AFs that represent a windfall to some manufacturers (e.g., the figure 6(a) monopolist) and a capricious loss to others who are severely disadvantaged relative to rivals located elsewhere in the U.S. and abroad (e.g., the figure 6(b) duopolist).

2.5 Firms are likely to respond asymmetrically to input price declines and input price increases.

FRR emphasize a key advantage of the time period they analyze is that it encompasses the period of the "shale boom," during which fracking technology enabled substantial deposits of hydrocarbons to be profitably extracted, significantly lowering natural gas and oil prices, and, in turn, the aggregate energy price variable constructed by FRR. These authors note correctly that their elasticities of sales response to energy prices are identified primarily from price decreases due to the shale boom and, thus, that the ability of those elasticities to forecast response to an energy price increase from cap-and-trade hinges on firms' responding symmetrically to price increases or decreases.

Asymmetric responses, however, have been found in many contexts. Standard econometric methods are available to test for possible asymmetric responses to price increases or decreases (e.g., Cameron et al. 1997), but were apparently not utilized in either study. An obvious reason for asymmetry is plant capacities. If a plant is operating at efficient capacity, then it has little short-run opportunity to expand production in response to an energy price decrease (e.g., due to the shale boom), but it would not be similarly constrained in reducing output due to higher energy prices caused by climate policy (see, e.g., Bushnell and Humber 2015). Thus, estimates of responses to input price decreases very likely underestimate responses to input price increases and represent another reason the FRR and GLM estimates likely understate the production response of California firms to cap-and-trade.[16]

3 Critiques of empirical modeling

3.1 Data limitations likely induce considerable measurement error that attenuates estimates of output responses to climate policy.

As already articulated in this report, data limitations hinder reliable inference about the magnitudes of production responses that can be expected in response to California's cap-and-trade program. In particular, poor proxies for industry output and industry long-run price expectations likely attenuate estimated production responses. Proxies for other variables that enter into the FRR and GLM estimating equations induce additional classical measurement error that further attenuates estimated responses.

Not only do FRR and GLM not observe firms' long-run price expectations that are relevant

[16]For example, suppose a firm had an annual production capacity of 100 units and in an initial equilibrium was producing 99 units. Assume that a given percent increase in energy prices would cause the firm to reduce output to 90 units. That same percent decrease in energy prices could cause at most a one unit increase in production.

to modeling firm responses to cap-and-trade, but they do not even observe the firm-level or industry-level energy prices they wish to employ as a proxy for price expectations. Instead, for domestic fuels prices, FRR and GLM compute industry-specific fuel shares from the quadrennial Manufacturing Energy Consumption Survey and combine these with state-level fuels prices from the U.S. Energy Information Administration's (EIA) State Energy Data System (SEDS).[17] The EIA admits SEDS prices are, themselves, estimates because "reliable data for state-level prices rarely exist, especially as a series that are consistent over a long period" (EIA 2014).[18] These aggregate price estimates are presumably noisy estimates of the contemporaneous fuel costs facing any particular plant. Moreover, energy-intensive firms may hedge fuel price risk with long-term contracts that would leave them unexposed to energy price changes in the short run.

FRR and GLM also assign establishments producing in more than one industry to a single NAICS classification for the purposes of aggregating production and imputing prices. FRR assign establishments to NAICS-6 industries according to the product that generates the greatest value for the firm across all years.[19] GLM do not articulate how they accomplish this aggregation. The aggregation of multi-product establishments induces further measurement error of energy prices because it imputes to at least a portion of the production at these plants an energy intensity, and, consequently, an energy cost that does not reflect the cost imputed to other firms producing

[17] FRR compute region and industry-specific fuel cost shares.

[18] EIA (2014) further explains:

> Estimates and assumptions are applied to fill data gaps and to maintain consistent definitions in the data series over time. SEDS incorporates the most consistent series and procedures possible for these estimates and assumptions. However, users should recognize the limitations imposed on the system due to changing and inadequate data sources. Estimates often are based on a variety of surrogate measures that are selected on the basis of availability, applicability as indicators, continuity over time, and consistency among the various energy commodities.
>
> It is also important to note that, even within a state, a single average price may have limited meaning in that it represents a consumption-weighted average over a whole state. For example, urban and rural electricity prices can vary significantly from a state's weighted average, and prices in one region of a state may differ from those in another because of access to less expensive hydroelectricity. Differences within a state may also be greater than differences among adjacent states. Thus, the principal value of the estimates in these tables lies in general comparisons among the states, interstate comparisons for a given year, and the analysis of trends over several years.

[19] NAICS-6, NAICS-4, and NAICS-2 refer to the North American Industry Classification System, the standard used by U.S. statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. economy. NAICS-2 refers to two-digit NAICS codes that classify firms by sector, NAICS-4 are four-digit codes defining industry groups, and NAICS-6 are six-digit codes defining individual industries. These codes are used by FRR and GLM to define industries and to categorize industries into groups and sectors.

in those industries, let alone the energy costs perceived by the establishments. This is likely to further attenuate the parameter estimates in FRR and GLM.

FRR also incorporate foreign energy prices in their international leakage modeling, assuming "the relevant foreign energy prices to consider are the prices in countries where imports originate and where exports are destined." Basic trade theory, however, suggests this assumption is wrong. In particular, the relevant prices to measure export competition are the prices faced by competing exporters, not the importers. Moreover, the focus on trading partners ignores the potential for import and export displacement by the trading partners of those trading partners. For instance, trade flows between the U.S. and a given trading partner may respond to price changes in a country that trades with a U.S. trading partner. But these price changes are ignored by FRR. Even accepting FRR's assumption about the relevant foreign prices, they are presumably estimated with error. In particular, FRR calculate industry-specific average electricity and natural gas prices for each export destination and import origin by aggregating across countries according to relevant trade volumes. This aggregation likely smooths potentially important price variation across countries that affects trade flows.

Because FRR are ultimately interested in emissions leakage resulting from changes in input costs to California production alone, they would ideally observe changes in California production and trade flows. But they do not observe quantities of goods that originate in or are destined for California. In their main analysis, then, they estimate changes in the values of national trade flows due to changes in energy prices. Their reliance upon national trade flows may induce additional downward bias in their estimates of import and export responsiveness of California production if California production is more trade exposed than national production, e.g., due to its proximity to the largest ports in the U.S. at Los Angeles and Long Beach (LALB). The LALB ports constitute the sixth largest ports by volume in the world. In a subset of their analysis, FRR restrict their attention to changes in shipments passing through California ports, which they admit is a "crude proxy" for changes in imports to California and exports from California plants. Indeed, it is. The ports receive 50% of all U.S. imports, and 77% of them are destined for markets beyond the region.

For example, sixty percent of goods imported to Chicago pass through LALB.[20] FRR acknowledge in Fowlie et al. (2016a) that "landlocked markets are sheltered from foreign competition." To the extent California production is more trade exposed than production elsewhere, e.g., because of port proximity, then either of their crude proxies for movements of California goods induces a downward bias in estimates of leakage due to California cap-and-trade.

3.2 Theory does not provide guidance as to the function that best relates production outcomes to energy price expectations (or energy prices).

Though theory predicts that higher input costs to California producers will induce reductions in in-state production, it does not define the precise nature of the relationship, and, therefore, provides little guidance to the econometrician who wishes to specify this relationship in an estimating equation. Therefore, as FRR write:

> To get any empirical traction, researchers need to make assumptions about the structure or functional form of these economic relationships. These choices can be fairly arbitrary, so it is important to evaluate how sensitive empirical results are to alternative plausible assumptions.

As Young and Kroeger (2015) assert, in areas where there is less certainty about methods, but also high expectations of transparency, researchers face a high burden of demonstrating broad robustness of model results. Neither FRR nor GLM meets this burden. FRR estimate 192 variations of a baseline model for each outcome, and report that their estimated elasticities, i.e., leakage rates, are sensitive to model specification. GLM vary select components of their baseline models, primarily the manner in which certain explanatory variables are defined and constructed, though parameter estimates, standard errors, and model fit statistics are not reported for any of their models.

In order to characterize model uncertainty, FRR present the distributions of elasticity estimates at each quartile of the distribution across all 192 models. But they present regression results (including parameter estimates, standard errors and model fit statistics) only for illustrative

[20]See, e.g., http://rtpscs.scag.ca.gov/Documents/2012/pfinal/SR/2012pfRTP_GoodsMovement.pdf

regressions. As Young and Kroeger (2015) noted, citing Ho et al. (2007), the reporting of only a curated set of model outputs generates "asymmetric information between analysts and readers. It is hard for readers to know if the reported results are powerfully robust to model specification, or simply an 'existence proof' that significant results can be found somewhere in the model space."

Throughout their analysis, FRR remain agnostic as to which model best describes the relationship between energy costs and production value. And they do not characterize the emissions leakage implications of those models that may be considered "best." Given the absence of theoretical guidance in modeling the production response to changes in input prices, and the apparent sensitivity of model outputs to specification, model uncertainty is considerable. FRR's attempt to characterize this uncertainty is noble, but it is largely ignored in FRR's subsequent simulations and in the ARB staff's proposal for allowance allocations, rendering it of limited usefulness. ARB staff calculates emissions leakage using median parameter estimates from the 192-model distribution of estimates without justification and without reporting the characteristics of the models yielding those median estimates. We don't know their functional forms, their model fit statistics, or their precision.

The agnosticism of FRR is surprising given established criteria with which to evaluate model fit. Of these, model cross validation is commonly preferred because it evaluates model performance on a hold-out sample of the data, implicitly penalizing models that over-fit sample data but perform poorly out of sample. Model cross validation could be used by both FRR and GLM to identify models that best fit the data. Subsequent analysis and AF determinations could be based upon these models that are presumed to better specify the relationship between energy input costs and production outcomes. At the least, FRR should report the specifications that inform their and ARB's subsequent analysis, and they and GLM should report model fit statistics.

It is beyond the scope of this review to consider all possible alternatives to the simple, static, and linear-in-logs models for which FRR and GLM report relevant results. Still, it is easy to imagine models that better comport with theory and the data than those they employ. The first of these would be models that incorporate the dynamics of firm adjustments. Though FRR report that a dynamic model is included among their 192 specifications in order to reflect the dynamic adjustment process, they report no results from any such models, apart from immersing them somewhere in the distribution of elasticities implied by the 192 models. GLM apparently estimate no dynamic models.

Instead, the models reported in FRR and GLM are static, relating production outcomes to contemporaneous prices without allowing past observations on dependent or independent variables to influence future values. In contrast to these static models, dynamic models are likely more fitting to the experimental setting. In particular, firms are likely to face costs of adjusting to changes in input costs because of inventories, contractual commitments, capital constraints, etc. Adjustments, therefore, are likely not to fully occur instantaneously. Dynamic models could account for these adjustment costs. A partial adjustment model, for instance, would allow FRR and GLM to estimate both a short-run and long-run response to input price changes. The abstraction from dynamics is surprising given that Fowlie et al. (2016a) explicitly emphasize dynamics, particularly changes to industry structure, in estimating the effects of output-based allowance allocations and alternative regimes on the cement industry. Comparing dynamic models to static models, they conclude:

> These two modeling frameworks predict substantively different welfare effects, thus highlighting the role of dynamic processes in determining the long-run welfare effects of these environmental policies.

Fowlie et al. (2016a) concede that "a dynamic model can aggravate the extent of leakage to unregulated areas by accelerating exit and retirement of regulated production units in the [regulated] market."

Given adjustment costs, it seems likely that the response to input price changes will be nonlinear and asymmetric. In particular, firms may not respond at all to small increases in input prices because the costs of adjusting production are greater than the avoided input costs, and they may respond differently to price increases vs. price decreases. Over a range of input price changes, then, no contemporaneous firm response may be observed. Beyond a minimum threshold, however, firms may adjust production in order to avoid higher costs. At the least, this suggests a piecewise linear response in price. However, if we imagine a distribution of firm adjustment costs across plants, even the piecewise linear function may insufficiently characterize the nonlinearity of firm response.

Third, both FRR and GLM might consider less parametric approaches to modeling heterogeneous responses. Both FRR and GLM include in their baseline estimating equations energy price interactions with energy intensity because "a given energy price increase should have a greater effect on the outcomes for energy intensive industries than for other industries" (GLM, p. 4) Though this likely is true, the reliance of FRR and GLM on this interaction term is seemingly mostly a matter of convenience. It affords them the ability to estimate industry-specific elasticities equipped with only two common parameter estimates (or three in the case of GLM's triple interaction of energy price, energy intensity, and industry group) and measures of industry energy intensity. It also allows them to claim to validate the ARB's determination of relevant industries as those that are EITE. Given the exclusion of other firm heterogeneity in their modeling, such claims of validation seem dubious.

Moreover, the use of an energy-intensity interaction with energy prices constitutes a parametric approach to estimating heterogeneous responses to energy price changes. In the parametric approach, it is important to correctly model the heterogeneity in responses. Less parametric approaches, that are less dependent upon modeling assumptions, include quantile regressions that separately estimate marginal effects from subsets of the data defined by energy intensity. Another alternative would include indicator variable interactions with energy prices where the indicators reflect categories of energy intensity. In fact, GLM adopt this approach while retaining energy price and energy intensity interactions. (They don't, however, report regression results for these or any of their estimated coefficients.) While these less-parametric approaches relax modeling assumptions, they do not directly yield industry-specific elasticity estimates, making them less convenient for FRR and GLM. Regardless, neither study reports sensitivity of parameter estimates to their specific modeling of industry heterogeneity.

Finally, in their limited specifications, GLM make several ad hoc modeling assumptions that at least warrant sensitivity analysis. For instance, they include indices of labor costs and demand growth that would be unnecessary if their identifying assumption were valid, i.e., energy price changes were as good as random conditional on the suite of fixed effects. We wonder whether model results are sensitive to the inclusion of these additional covariates, and if they are, then what other potential omitted variables may bias their results.[21]

GLM also choose to model the dependency of firm production on energy prices of potential rivals in an ad hoc fashion described on page 8. In particular, they attempt to identify competing firms to any given plant as those located in states for which about 10 percent of production is located within 250, 500, or 1,000 miles of 10 percent of production in the state in which the given plant is located. Again, there is no indication of how sensitive parameter estimates are to these assumptions. In particular, GLM report elasticity estimates from models defining competitors to be located within 250 and 500 miles. This seems a restrictive definition of competition among manufacturers, and GLM do not report robustness to more generous definitions. Moreover, because competitor energy prices are defined as the simple average of energy prices across states with potential competitors, potentially important interstate variation is lost to aggregation. What variation remains is likely absorbed by region-year fixed effects. These controls become effectively worthless.

3.3 There is a tradeoff between avoiding omitted variables bias and avoiding "errors in variables" bias.

Griliches and Hausman (1986) articulated a tradeoff between omitted variables bias and "errors in variables" or measurement-error bias when measurement error is a concern. As previously asserted in this report, measurement error is likely to be a serious problem in the context of the FRR and GLM analyses chiefly because contemporaneous energy prices, themselves noisily measured, proxy for long-run energy price expectations, but also because proxies for output changes are attenuated. The well-known analytical result of Griliches and Hausman (1986) shows that if an independent variable is measured with error, then estimates of the relationship between the dependent and independent variable are inconsistent, and under certain (but general) assumptions, the coefficient estimate is attenuated towards zero (McKinnish 2008). As McKinnish (2008) notes:

In particular, we expect many outcomes to respond differently to short-term and

[21] We note that the suite of fixed effects GLM include is not as robust as the fixed effects Linn (2013a) and Linn (2013b) asserted were sufficient to generate exogenous variation in energy prices.

long-term variation in conditions. This differential effect of long-term and short-term variation can generate the same bias as "true" measurement error.

Thus, contemporaneous energy prices upon which FRR and GLM rely in their estimation mis-measure price expectations, and, likely induce attenuation in the FRR and GLM leakage estimates from panel data models.

The measurement error problem is especially severe in panel data models because they eliminate much of the information about the true independent variable contained in the proxy when they eliminate the time or industry fixed effects that might otherwise cause omitted variables bias. What is left following either differencing or the within transformation is primarily the noise component of the proxy variable. FRR defend their reliance upon contemporaneous energy prices by appealing to a sustained period of low energy prices during the shale energy boom. The implication is that this secular trend in energy prices was likely reflected in energy price expectations. As we have already asserted, however, energy markets were not expected to sustain low prices for such a prolonged period. Moreover, in implementing fixed effects for NAICS-6, and NAICS-2 interacted with year, FRR sweep away the secular trend in energy prices that is expected to induce production responses. They are left to identify the production response from only annual variation in energy prices within sectors, which reflects more idiosyncratic variation and less secular trend than FRR's appeals to the shale boom secular decline would suggest. At the same time, because the persistent decline in energy prices due to the shale boom coincided with the Great Recession, and because both likely had heterogeneous impacts across industries, it is conceivable FRR's estimation nevertheless suffers from omitted variables bias, i.e., from time-varying intra-NAICS-2, or industry-specific, effects.[22]

GLM implement a suite of fixed effects that is likely to even better control for omitted variables bias than FRR and yet leave even less secular variation with which to identify production responses. GLM include, in addition to industry and year fixed effects, region and year interactions, as well as industry and year interactions. They are left to identify production responses off of variation in energy prices received by firms within industries and regions holding constant region

[22]NAICS-4 and year interactions would mitigate concern about omitted variables bias. FRR apparently refrain from doing so in their illustrative regressions, at least, in order to avoid over-saturating the model with fixed effects.

and sector changes in prices. This energy price variation is likely to be very idiosyncratic and unlikely to drive producer responses. If production were thought to respond in the short run to even small deviations in prices, then this problem would be mitigated. But the dynamic adjustment that is likely to attend only persistent changes in long-run price expectations is unlikely to permit such short-run responses.

Finally, it is important to note that, when using panel data methods for causal inference, researchers often emphasize a robust suite of fixed effects in order to eliminate potential omitted variables bias and to reflect conservatism in resulting parameter estimates. This is most appropriate when the signs of coefficients and their statistical significance, i.e., their difference from zero at standard levels of scientific certainty, are of interest and when outcomes are likely to reflect responses to idiosyncratic, rather than secular, changes in explanatory variables. In some settings, it is sufficient to document that a statistically significant causal relationship exists between an independent variable and an outcome of interest.

In the present setting, however, the *magnitudes* of the parameter estimates are of critical importance because they determine the magnitudes of expected emissions leakage and provide the basis for setting AFs. Signs (of coefficients) and significance are insufficient for this task. Consequently, conservatism in model specification, i.e., by wiping away secular variation in the independent variable in order to avoid endogeneity, does *not* yield conservatism in leakage mitigation. Rather, it does the opposite by potentially biasing estimates of output responses towards zero. An alternative approach would rely less on fixed effects, accepting greater risk of omitted variables bias in exchange for less attenuation bias. The tradeoff between omitted variables bias and attenuation bias is a costly one that renders convincing estimates of leakage risks nearly impossible to produce without information about how energy price expectations evolved over the sample period.

3.4 The variance of estimated parameters is largely ignored in each study, though it likely renders estimated industry-specific leakage rates statistically indistinguishable.

FRR relax modeling assumptions by specifying 192 unique estimating equations, and they implicitly rely upon motivations for model averaging by selecting parameters from the median of their parameter distributions. In this way, they effectively account for model uncertainty in their estimation, though this uncertainty is ignored in the ARB staff's subsequent analysis. Regardless, there is very little attention in FRR to the variance of estimates in their report. They report standard errors for the select "illustrative" models they estimate, but do not discuss uncertainty for the models generating median parameter estimates upon which they and ARB staff rely in subsequent analysis.

More egregiously, GLM do not report standard errors for any of their models anywhere. Tables in GLM that report model results include asterisks on some parameter estimates, presumably signifying estimates that are statistically different from zero at the 10% level, in keeping with convention. But as previously asserted in this report, "different from zero" is not good enough for the ARB's purposes, as it intends to generate distinct industry AFs. The magnitudes of estimated firm and industry responses are of paramount importance in this context, as is the precision of those estimated magnitudes.[23]

In their figure 5, FRR depict industry-specific elasticity estimates and standard-error-based 95% confidence intervals for one of their 192 specifications. They array these elasticities according to energy intensity and demonstrate that responses become more elastic at greater energy intensities. This figure, however, is misleading for several reasons. First, it reports confidence intervals for only one of several models. The reader is not told which model generated the parameter estimates or whether other models generated larger confidence intervals implying less precision and more uncertainty in parameter estimates. Second, the figure engenders an impression that

[23] Abadie et al. (2014) demonstrate that uncertainty persists in estimated causal effects due to conditional random assignment even if sampling uncertainty is diminished as the sample size approaches the size of the population of interest. That is, even if FRR and GLM observe the entire populations of interest, they do not observe the entire populations in all potential outcomes. Instead, they have random sampling of potential outcomes. Thus, uncertainty persists and the frequentist standard errors for causal estimates are generally appropriate.

the elasticity estimates are each unique, and, therefore, the pattern of increasing elasticity in energy intensity validates subsequent analysis. This pattern, however, is not a consequence of unique, industry-specific elasticity estimates. Rather, it is a function of FRR's baseline model specification, namely the interaction of energy prices with energy intensities and the common negative relationship between revenues and this price-and-energy-intensity interaction across firms. The trends in figure 5 are a consequence of this modeling assumption rather than a validation of it. Finally, the confidence intervals depicted in the figure, suggest there is no statistically significant difference in the production, import, or export elasticities of the vast majority of industries they study.

GLM do not report any regression results at all, contending that "because there are so many parameters in the model, it is difficult to interpret the estimates of individual parameters" (GLM p. 16). They instead report only industry-specific elasticity estimates derived from these individual parameter estimates. They present them as though they were estimated with certainty. Likewise, their depiction of elasticities by cost share in figures 1 and 2 omit any indication of sampling uncertainty embodied in the point estimates. These elasticity estimates are uninformative without knowledge of the error structure around them. Like FRR, they conclude from figures 1 and 2 that larger elasticities among more energy intensive industries "is consistent with expectation and supports the validity of the modeling approach." Again, however, this result is a consequence of the modeling approach, and, therefore, is of little value in validating the approach.

3.5 The exclusion of non-EITE industries forecloses an opportunity to validate the empirical approaches.

Both FRR and GLM limit their statistical analysis to EITE industries that ARB has determined are eligible for allowances. While this may seem sensible, it, in fact, forecloses an opportunity to validate the empirical methods of FRR and GLM and the industry classifications of ARB. If ARB has correctly classified industries as EITE, then the responsiveness of non-EITE industry production and trade flows should be distinct from the responses of EITE industries. In particular, they should be distinctly smaller than the responses of the EITE because (a) energy prices impact total costs to only a nominal degree, and (b) there is less opportunity for production to avoid high energy input costs through leakage. In fact, this is such a prime opportunity and obvious way to validate the ARB classification and the precision of parameter estimates in GLM and FRR, that ARB staff should conduct falsification tests and robustness checks on these parameter estimates using non-EITE industries. The fact that these robustness checks have not been conducted raises questions about the validity of the FRR, GLM and ARB methods.

4 Assessment of ARB staff analysis

Since submission of the FRR, GLM, and Hamilton et al. studies, ARB staff has prepared multiple reports outlining its planned recommendations to the ARB regarding use of the studies to set AFs beginning in 2020. In "Appendix E Staff Report: Initial Statement of Reasons," released August 2, 2016, Staff proposes that beginning in 2020, allowances allocated to industries for transition assistance will be eliminated, and AFs will depend entirely on "assistance necessary to prevent emission leakage" (p. 3). In the report, staff endorses the FRR and GLM studies, noting that "these two studies complement each other to provide a complete picture of emission leakage potential for most manufacturing sectors (p. 4)."

Staff proposes a framework whereby percent assistance factors will be assigned for each manufacturing industry by summing an international leakage mitigation AF based upon the FRR study and a domestic leakage mitigation AF based upon the GLM study. The international AF is based on the "international market transfer" (IMT) metric proposed by FRR, which is defined as the fraction of the decrease in domestic production caused by higher energy costs that is offset by an increase in international production.[24] Staff indicates on p. 11 that the IMT would also be "informed" by Hamilton et al. (2016). However, in the follow-up report dated October 14, 2016 wherein tentative recommendations on industry-specific AF are made, no use is made of the results

[24]Imprecision of the parameter estimates and lack of transparency regarding this imprecision has been a major theme of this review. It is worth noting that IMT is constructed as a ratio involving three random variables (see FRR p. 39), the domestic, export, and import elasticities with respect to the energy price. It is a well-known result from statistics that confidence intervals for ratios of random variables can be very wide (Fieller 1954, Scheffe 1970), even if the random variables themselves exhibit only modest dispersion. FRR (p. 38) appear to be aware of this problem, and point to it in cautioning policy makers regarding constructing industry-specific IMT from their estimates.

in Hamilton et al. 2016, and the study is not mentioned at all.

The staff rejects a similar metric to IMT for the domestic industry due to the assumption that production declines in California will be met one-for-one by increases in production elsewhere in the U.S. Such an assumption would result in 100% AF on the domestic market under the FRR approach. Instead, staff recommends setting the domestic AF so that domestic sales by California firms do not drop below a threshold level. In its October 14 report staff sets this threshold production decline (known as "domestic drop" or DD) at 7%. Thus, allowances would be offered to prevent DD from exceeding 7%.

In Appendix E, staff sets forth the proposition that each of the leakage studies has erred on the side of conservatism such that the leakage risks measured by FRR and GLM represent "the upper bound of the manufacturing sector's potential emissions leakage risk levels" (p. 6). As noted, we strenuously disagree with this characterization. We return to this very important point in the next section of our report. Staff also announced in Appendix E a plan to introduce further conservatism into the setting of AFs. Specifically staff proposes to create additional versions of IMT and DD based on its own analysis of the FRR and GLM results. AFs would, then, be based upon the computations of IMT and DD that were most favorable to the impacted industries, i.e., that yielded the highest AFs. Although ARB staff emphasizes this point repeatedly in Appendix E, the plan is abandoned in its October 14 report entitled "Cap-and-Trade Regulation Assistance Factor Calculation: Informal Staff Proposal." The report instead proposes to average the AFs that emerge from the various metrics developed by FRR, GLM and ARB staff.[25]

In the October 14 document, ARB staff derives new estimates of IMT and DD using regression analysis of the results provided by FRR and GLM, respectively. Staff also describes various adjustments it makes to the FRR and GLM results in order to eliminate outcomes from those studies that staff determines are anomalous or counter to economic intuition. Staff creates a new measure of IMT by regressing the FRR estimates of IMT (based upon the median parameter estimates from the distribution generated by FRR's 192 regressions) on measures of industries' trade exposure and energy intensity.

[25] The October 14 document was subsequently revised on October 21, with changes limited to tables 3-6 from the October 14 document.

The analysis by ARB staff is problematic for several reasons. First, for the measure of IMT adapted directly from the FRR study (called "raw IMT"), staff adopted and utilized without explanation or justification the so called "P50" or median FRR estimates of domestic, export, and import elasticities. It is improper for staff to choose the midpoints of the respective parameter distributions and use them to compute IMT as if each of these components of the IMT were estimated with certainty. This is not something that FRR did in their report, and, in fact, they issued strong cautions again doing precisely what ARB staff does:

> The natural next step, from the perspective of a policy maker looking to assess leakage risk and target leakage mitigation measures, is to translate these responsiveness measures to corresponding measures of market transfer and associated emissions leakage. However, pushing on to this next step amounts to pushing up against the limits of available data. One complication is that calibrating the measures of leakage risk implied by the theory requires dividing one noisy estimate by another (p. 38).[26]

FRR further caution:

> Note that these industry-specific transfer rates are constructed as a ratio of our imprecise elasticity estimates. A ratio of noisy numbers can be very noisy; our industry-specific estimates of market transfer rates are sensitive to changes in how the underlying estimating equations are specified (p. 39).

In evaluating this decision by staff, it is worth summarizing some of the points we made in detail in the prior section. We know nothing about the statistical properties of the regression models that generated the median elasticities used to compute IMT or even their functional forms or the variables included in those models.[27] No attempt is ever made by FRR or ARB staff to show that the models yielding these median elasticities have better statistical properties than any of the other models they estimate. The fact that these models yield elasticities that happen to fall at the

[26] In an Excel data file delivered to ARB, FRR do compute IMT based on estimates at the 25th, 50th, and 75th percentile of the distributions from their 192 model specifications for domestic, import, and export sales elasticities. This spreadsheet is unaccompanied by the authors' analysis of these values.

[27] We assume that the median elasticities for domestic, export, and import revenues emanated from different regression models. There is no reason to believe that a single model would have produced the median estimate for each elasticity.

midpoint among the 192 estimates provides little justification for their use. Moreover, the sampling uncertainty of these parameter estimates is nowhere reported in FRR, so we cannot determine to what extent industry-specific IMTs are statistically distinguishable. FRR, however, admit:

> Given the noisiness of these estimates, we cannot estimate the transfer rate for any given industry with any degree of confidence (p. 39).

The "raw IMT" computed from median parameter estimates in FRR become the dependent variable in Staff regressions that produce IMT estimates as a function of energy intensity and trade exposure. ARB Staff first adjusts negative industry values of the raw IMT to zero and raw IMT greater than one to precisely 1.0. We can appreciate why ARB staff felt compelled to change these estimates of IMT because they plainly conflict with theory. Their existence is yet another reflection of the imprecision that permeates these empirical studies. We agree that negative IMT or IMT in excess of 1.0 are nonsensical and problematic, but we disagree with the cure employed by ARB staff, which is effectively to sweep the problem away in an ad hoc manner. Just as these obviously anomalous results reflect imprecision in IMT estimates, surely so too are many of the IMT that lie between zero and 1.0 also highly inaccurate; it is just less obvious to the "naked eye." They perhaps warrant similar adjustments, though it is less obvious in which direction and by which magnitude these estimates are also in error. Instead of seeing these types of results as ominous warning signs that the studies it commissioned generated imprecise and potentially inconsistent estimates of IMT that are ill-suited for determining AFs and the attendant sizable transfers, ARB staff simply made ad hoc adjustments to the results.

Unlike FRR, GLM report a preferred model based upon their equation (2), but they provide no evidence to justify their preference for it. ARB Staff adopt the results of this model for two of the three dependent variables estimated by GLM--sales revenue and value added--to compute estimates of DD, ignoring (appropriately in our view) the third--employment. Then they regress the DD estimates from GLM on industry energy intensity. Just as FRR generated counter intuitive parameter estimates, so, too, did GLM estimate positive values for some DDs. Staff adjusted these estimates downward to match those of industries with similar capital intensities, effectively imposing heterogeneous responses across some industries and overriding with staff judgment select outcomes of the econometric modeling it commissioned. Again, this constitutes an ad hoc adjustment necessary to rationalize imprecise estimates of production responses to energy prices. That these adjustments are necessary belies the fact that positive DDs are likely no further from the truth than many of the negative DD estimates that staff simply leaves alone because they are less obviously wrong.[28]

Staff's decision to construct additional measures of IMT and DD using regression analysis is curious in many dimensions. Most importantly staff's regression-based estimates provide no new information and do nothing to overcome the limitations of the empirical studies in estimating emissions leakage. Combined with the ad hoc adjustments to the FRR and GLM results, they have only the effects of "smoothing" the FRR and GLM estimates of IMTs and DDs in order to eliminate outliers and permitting out-of-sample calculations of IMTs and DDs. It does not improve inference or the precision of IMT and DD estimates. It simply cannot improve precision because it contributes no new information to the inference problem. For example, staff regresses estimated DD for each six-digit NAICS industry from the GLM preferred model specification on the log of energy intensity for those industries. Energy intensity interacted with energy price represented an explanatory variable in GLM's preferred model, so the DD estimates that constitute dependent variables in the staff regressions are by construction functions of energy intensity. Unsurprisingly, then, energy intensity appears to explain DDs.[29] In sum, the staff regressions reduce the range of estimates of DD and IMT across industries, but they are based on no new economic information, and there is no reason to believe that these fitted DD and IMT (what staff calls "regressed" estimates) are any closer to the true values than the original estimates.

In the end, ARB staff has two measures of IMT, one that uses the median elasticity estimates from FRR (subject to staff adjustments), and the other from their regression of the

[28] In reality positive values for DD are not counter intuitive once one remembers that sales revenue and value added are not the correct variables to be utilizing in leakage analysis because they conflate the desired pure output response with a price response. If demand is sufficiently price inelastic, the percent price increase will exceed the percent output decrease due to higher energy costs, causing both sales revenue and value added to increase. This point was raised originally by Levy (2016) in the Brattle Group comment on the leakage studies.

[29] For instance, if the demand drop for a given industry i, DD_i, is small relative to the industry's energy intensity, then the regression of DD on energy intensity will generate for this industry a positive residual, $\hat{DD}_i > DD_i$, where $\hat{DD}_i$ is the DD predicted by the regression equation. Similarly, an industry with a large DD relative to its energy intensity will generate a negative residual in the regression because the predicted value from the regression is less than the DD from the GLM model.

staff-adjusted IMT on trade exposure and energy intensity. Staff averages the two IMTs to produce its recommendation of AF for international leakages. It has four estimates of DD—two from the GLM preferred model for value added and sales revenue (subject to adjustments by staff) and two from the regressions of these staff-adjusted DD estimates on industry energy intensity. The AF that yields a 7% output reduction (DD) is computed for each estimate, and these AFs are then averaged to obtain staff's recommended domestic AF. This averaging across measures is not what ARB staff indicated it would propose in August when Appendix E was released.

After all of the manipulations, the proposed AFs are largely determined by industries' trade exposure and energy intensity. Energy intensity is the key variable introducing heterogeneity in the elasticity estimates generated by FRR and GLM, and it assumes even greater prominence following the ARB Staff adjustments of the FRR and GLM results and their regressions. Energy intensity is an important factor from a cost-of-production perspective because it indicates the magnitude of increase in marginal production costs that will occur due to higher energy costs from cap and trade.

Trade exposure is also an important factor, but it provides a very incomplete picture of the competition faced by California manufacturing industries. A California firm that competes primarily in the international market with its major competitors located abroad will be given an AF that attempts to reflect this competition. A California firm that operates exclusively in the domestic market with competitors in the U.S. receives no comparable assistance, unless its DD exceeds the 7% threshold.

Finally, at the risk of stating the obvious, we nonetheless feel compelled to emphasize that none of the staff adjustments do anything to introduce into the AF-setting procedure the key elements that are missing from the GLM and FRR statistical analyses, such as demand-side and competition factors, as discussed earlier in this report.

5 Are the FRR, GLM, and ARB staff analyses "conservative"?

FRR, GLM, and ARB staff repeatedly assert that their analyses are conservative in the sense of favoring industry and yielding upper bounds on estimates of emissions leakages. If these assertions were correct, then proposed AFs would be greater than those necessary to satisfy the legislative mandate to minimize emissions leakage to the extent feasible. They promote the belief that California bears little risk in proceeding with the proposed AFs even if the industry-specific leakage rates are imprecisely estimated, which we contend and FRR admit they are. ARB staff aptly summarize this argument in Appendix E (p. 6):

> Each of the three studies makes conservative assumptions that result in leakage risk assessments at the upper bound of the manufacturing sector's potential emissions leakage risk levels.

In Appendix E, staff proposes to invoke additional conservatism by adopting the methodology for both IMT and DD that yields the highest AF.

The claims of conservatism in the FRR and GLM studies derive from the following:

- The GLM study and ARB staff assume a one for one tradeoff between California and other US production. In reality, outside production may not fully supplant California production, e.g., due to price effects.
- GLM estimate impacts in the long run that are much smaller than their short-run impacts, and, yet, ARB Staff utilize the relatively larger short-run estimates that imply larger AFs.
- In FRR, use of national totals for exports and imports overstates IMT. Decreased California exports may be replaced by expanded exports from U.S. firms outside California. Moreover, reductions in California exports may not be fully replaced by foreign producers due to price effects. Similarly, increased imports into California could reflect reallocations of output by exporting firms instead of expanded production.
- GLM assume a higher compliance cost than the 2016 emissions allowance reserve price in their simulations of industry leakage. Thus, they estimate leakage from a policy that raises energy input costs by more than they are likely to rise due to cap and trade. This begets a greater production response that implies greater AFs than are expected to be necessary.

Indeed, one-for-one offsets of reduced California production by domestic or international competitors may not occur for the reasons FRR and GLM enumerate. And while the ARB's reliance upon the relatively larger short-run elasticity estimates of GLM may favor conservatism, it is nevertheless warranted by the likely inconsistency of GLM's long-run elasticity estimates and by the incompatibility of the relative magnitudes of their short-run and long-run impacts with production theory as we and FRR understand it.[30] That is, the long-run estimates should exhibit greater responsiveness to input costs than short-run estimates. That they don't indicates a problem with the empirical strategy and should not be interpreted as imparting any particular confidence to the short-run estimates.

A determination of whether the emissions leakage estimated by FRR and GLM constitutes an upper bound, and whether the ARB staff's AFs err on the side of generosity to industry, requires that potential sources of conservatism in their analyses are weighed against factors that may cause estimated leakage rates to be too low. Discussion of these several factors is conspicuously absent from the FRR, GLM and ARB staff reports. Although we have enumerated these points throughout our review, we reiterate them given the importance of the "conservatism" claim.

First, because of California's coastal location and proximity to ports, it is very likely that exports from California and imports to California are more price sensitive (elastic) than are exports and imports for the country as a whole. Moreover, this omission in the FRR study is likely to selectively bias the leakage estimates for industries with high weight-to-value ratios that dominate ocean freight trade. If so, import and export elasticities estimated by FRR using national data would understate the elasticities facing California producers and introduce selective bias in estimates for particular industries.

Second, and importantly, California production has lower emissions than production elsewhere in the U.S. and most anywhere around the world. This is due, in part, to California's clean-energy policies, and, particularly, its clean electricity-generating fleet. In contrast, China,

[30] GLM even concede that their long-run estimates are unreliable, reporting that "for statistical reasons we suggest caution when interpreting the long-run results for individual industries" (p. 2). Though they do not articulate what are the statistical reasons, they do reiterate that "the industry-specific results suggest caution with using the long-run results for individual industries."

the largest exporter to the U.S., emits five times more carbon per unit of output than the U.S.[31] Despite general agreement that California production is less emissions-intensive than production most anywhere else, FRR and GLM assume leaked production is just as clean. Higher emissions intensities outside of California exacerbate emissions leakage and cause ARB's leakage estimates and AFs to be biased down, rather than up. Moreover, California industries that trade predominantly with "dirty" regions may result in leakage outcomes that are selectively biased relative to industries that trade predominantly with "clean" regions, for instance the EU.

Third, neither FRR nor GLM estimate impacts on physical output, and instead rely upon sales revenue (FRR and GLM) and value added (GLM) to infer IMTs and DDs. Both measures conflate output and price changes, which move in opposite directions. Because both sets of authors rely on national data, it is a virtual certainty that their dependent variables involve both price and quantity effects in response to energy price changes. Thus, they will measure less impact than if they were correctly measuring output due to the offsetting price impacts. As noted originally by Levy (2016), the net impact on sales revenue of an energy price increase will be positive if demand at the national level is sufficiently price inelastic. Moreover, given California firms' generally small national market share for manufacturing (e.g., Bernstein et al. 2016), an increase in energy costs that is restricted to California will typically not generate any significant offsetting price effect. These price effects, which are captured in the national data, mute the output response being measured by GLM and FRR. Thus, use of national data and sales revenue instead of output introduces two biases against conservatism in the estimated elasticities, one from conflating price and output effects, the other from understating the output effect if prices are held constant, as is likely for a California-only energy input price increase.

Fourth, by relying upon contemporaneous energy prices to proxy for the price expectations that are relevant to firm production decisions in the short run and long run and that change in response to cap and trade, both FRR and GLM are likely to underestimate leakage rates because

[31] China's imports into the U.S. were valued at $467 billion in 2015. Based on data released by the International Energy Agency (2009), China was the world's fifth "dirtiest" producer based on GDP in U.S. dollars per ton of carbon emissions of $435. By comparison the U.S. had GDP of $2,291 per ton of carbon emissions. Thus, on average a dollar of production in China emitted five times the carbon as a dollar of production in the U.S. The European Union, the second largest exporter to the U.S. is somewhat cleaner, with a GDP/ton carbon emissions ratio of $3,712. Canada is the third largest partner and has an emissions ratio almost identical to the U.S.

measurement error attenuates their estimates to zero. As McKinnish (2008) concludes:

> Fixed-effects and first-differences models are extremely popular because the relationship of interest is often confounded by unobserved heterogeneity in the cross-section. Unfortunately, if the independent variable is an imprecise measure of the relevant factor, coefficient estimates from these models can be severely attenuated towards zero. The time series variation that remains after removing fixed effects often largely reflects idiosyncratic changes in the independent variable that have little influence on the decision of interest.

In short, data limitations, and the choice by FRR and GLM to accept greater risk of attenuation bias in order to mitigate risk of omitted variables bias most assuredly render their estimates of leakage and ARB's AFs too low, rather than too high.

Fifth, and relatedly, the shale-induced secular decline in energy prices upon which FRR in particular rely to identify output responses coincided with the Great Recession and the ensuing sluggish recovery. These macroeconomic conditions would be a drag on output in most manufacturing industries, though it certainly affected some industries more than others. Industry-specific trends would induce omitted variable bias in FRR. The more robust suite of fixed effects in GLM reduce concerns about omitted variables bias, though, as they admit, do not necessarily deliver "as good as random" energy price variation. And, the robustness of their estimates to industry-specific trends comes at the expense of secular variation over which to identify firm production responses.

Sixth, as FRR note, they identify sales revenue effects off energy price changes that are primarily negative, and in order to apply those effects to price increases due to cap and trade, firm response to higher vs. lower energy prices needs to be symmetric. Capacity constraints are an important reason (especially for short-run analyses) why responses may not be symmetric. Firms that could not increase production from lower gas prices due to the shale boom because of capacity constraints can nevertheless reduce production due to higher energy costs.

Finally, the FRR and GLM studies all rely upon marginal analysis. The elasticities predict smooth changes in sales revenue as a function of energy costs. They cannot predict drastic output changes from firms closing plants, relocating production to states or countries with lower energy costs, or exiting the business entirely. To the extent such responses occur due to California firms facing a chronic energy cost disadvantage, they are not captured by either study and represent yet another bias against conservatism.

6 Conclusion

A thorough review of the studies conducted by FRR and GLM and of the ARB staff proposal for post-2020 AFs leads us to conclude that the staff-proposed AFs are likely insufficient to mitigate emissions leakage and may not satisfy the ARB's statutory obligation to mitigate leakage "to the extent feasible." Moreover, the staff-proposed, industry-specific AFs would assuredly create "winners" and "losers" among the regulated firms; firms with similar emissions leakage risks may receive different AFs, while firms with very different leakage risks may receive similar AFs. On balance, given the downward bias of leakage estimates, most or all regulated firms would be losers relative to the protections envisioned by the legislature when it mandated mitigation of leakage risks. We reach these conclusions chiefly because the methods employed by FRR and GLM inadequately account for heterogeneity in industry responses to input costs and make a number of assumptions and modeling decisions that are likely to produce estimates of industry responses to cap and trade that are too low, i.e. biased toward zero. Moreover, the estimates generated by the empirical modeling of FRR and GLM are highly uncertain, or noisy, and, therefore of limited use in assigning industry-specific AFs. The limitations of the panel-econometric approach employed by FRR and GLM are perhaps highlighted by the ARB staff's ad hoc changes to estimates produced by the FRR and GLM empirical models that do not comport with theory or their intuition. However, the same uncertainties and methodological limitations that generated the unintuitive results ARB staff are compelled to edit are also reflected in other estimates that ARB staff leave alone.

The research teams that prepared the FRR and GLM reports are skilled and hold relevant expertise. And yet, the task they were handed is an extremely difficult one given available data. For the most part, we would characterize their efforts as valiant ones that came up short. However, we are struck by their repeated and concerted efforts to characterize their empirical estimates of firm production responses as "upper bounds" on responsiveness that yield AFs that err on the side of generosity to industry. This is despite a number of assumptions, modeling decisions, and data limitations that almost surely render their estimates of emissions leakage too low rather than too high. The implications of these sources of downward bias in their estimates of firm responsiveness are never mentioned. Further, while FRR acknowledge modeling uncertainty, neither FRR nor GLM accounts for the considerable sampling uncertainty in their estimates or derivative calculations. Neither does the ARB staff, which proceeds to employ FRR and GLM results in assigning industry-specific AFs even though FRR expressly caution against doing so, saying, "Given the noisiness of these estimates, we cannot estimate the transfer rate for any given industry with any degree of confidence" (p. 39). Greater transparency in model performance and model outputs is also warranted in each study.

The limitations of these studies are so considerable, that we must reluctantly ponder whether "no number is better than some number," as Hausman did in 1994 and 2012 concerning contingent valuation of non-marketed goods (Hausman 2012, Diamond and Hausman 1994. Hausman asserted that the test of whether some number was better than no number was whether policy making was improved with "some number" relative to policy making without "the number." In the present setting in which the ARB must determine how to mitigate emissions leakage, it seems possible that no number from these studies may yield better policy outcomes than the numbers generated by FRR and GLM. This is perhaps the case because the ARB would be compelled to approach emissions leakage in a truly conservative manner, e.g., by continuing to allocate sufficient allowances so as to avoid imposing upon California firms considerable disadvantage.

More so, however, the ARB may consider an alternative approach to estimating leakage that accounts for industry heterogeneity. Though we did not undertake a review of Hamilton et al., our familiarity with the study leads us to believe that it may point a way forward for the ARB that would overcome some of the limitations of FRR and GLM. In particular, the panel econometric approach of FRR and GLM cannot account for the market structures of heterogeneous industries. But industry-specific analyses such as those performed by Hamilton et al. could account for these differences and yield theoretically consistent, industry-specific leakage estimates. The success of such studies would depend upon the reliability of demand and supply elasticity estimates found in the extant literature or estimated from available data. FRR, themselves, highlight the potential for industry-specific structural and dynamic modeling of leakage risks in the cement industry in Fowlie et al. (2016a), while Bushnell and Humber (2015) assess production responses of the nitrogenous fertilizer industry. Such industry-specific empirical approaches hold more merit than the reduced-form models of FRR and GLM that largely impose industry homogeneity.

We further conclude that it is very challenging to get leakage mitigation right. Given data limitations, it is difficult to imagine how one might arrive at precise estimates of industry-specific emissions leakage risks. And yet, the success of regional cap and trade programs depend upon it. Without it, production and emissions leave the regulated jurisdiction only to be replaced by production and emissions outside the region. The regulated economy suffers and to little or no environmental benefit. In fact, given California's relatively clean fuel mix, absent effective leakage mitigation, worldwide emissions could increase as a consequence of California's cap and trade program. While a binding worldwide commitment to emissions reductions could mitigate such emissions leakage risk, the explicit or implicit pricing of GHG emissions elsewhere around the world seems less certain at the close of 2016 than it did only a year earlier.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge, "Finite Population Causal Standard Errors," 2014. Online at http://www.nber.org/papers/w20325. Last visited December 2016.

Ailworth, Erin, "Despite Glut of Oil, Energy Firms Struggle to Turn Off the Tap," *The Wall Street Journal*, 2015, *August 16*. Online at http://www.wsj.com/articles/despite-glut-of-oil-energy-firms-struggle-to-turn-off-the-tap-1438904654. Last visited December 2016.

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly harmless econometrics: An empiricist's companion*, Cambridge, MA: Princeton University Press, 2008.

Bushnell, James and Jacob Humber, "Rethinking Trade Exposure: The Incidence of Environmental Charges in the Nitrogenous Fertilizer Industry," 2015. Online at http://economics.ucdavis.edu/events/papers/1028Humber.pdf. Last visited December 2016.

Cameron, A Colin, Severin Borenstein, and Richard J Gilbert, "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes?," *Quarterly Journal of Economics*, 1997, *112* (1), 305–39.

Davis, Lucas W and Lutz Kilian, "Estimating the effect of a gasoline tax on carbon emissions," *Journal of Applied Econometrics*, 2011, *26* (7), 1187–1214.

Denning, Liam, "Power Investing," *The Wall Street Journal*, 2010, *September 13*. Online at http://www.wsj.com/articles/SB10001424052748703846604575447762301637550. Last visited December 2016.

Diamond, Peter A and Jerry A Hausman, "Contingent Valuation: Is Some Number Better than No Number?," *Journal of Economic Perspectives*, 1994, *8* (4), 45–64.

Economist, The, "Better Out Than In," *The Economist*, 2013, *March 2*. Online at http://www.economist.com/news/leaders/

21572769-if-barack-obama-wants-cleaner-world-and-richer-america-he-should-allow-natural-gas. Last visited December 2016.

Energy Information Administration (EIA)

Energy Information Administration (EIA), "Annual Energy Outlook 2007," February 2007. Online at http://www.eia.gov/oiaf/archive/aeo07/pdf/0383(2007).pdf. Last visited December 2016.

Energy Information Administration (EIA)

_ , "Annual Energy Outlook 2010," April 2010. Online at http://www.eia.gov/oiaf/aeo/pdf/0383(2010).pdf. Last visited December 2016.

Energy Information Administration (EIA)

_ , "Annual Energy Outlook 2012," June 2012. Online at http://www.eia.gov/outlooks/aeo/pdf/0383(2012).pdf. Last visited December 2016.

Energy Information Administration (EIA)

_ , "Today in Energy: Projected natural gas prices depend on shale gas resource economics," August 2012. Online at https://www.eia.gov/todayinenergy/detail.php?id=7710. Last visited December 2016.

Energy Information Administration (EIA)

_ , "State Energy Data System 2014 Price and Expenditure Technical Notes," August 2014. Online at http://www.eia.gov/state/seds/sep_prices/notes/pr_technotes.pdf. Last visited December 2016.

Fieller, Edgar C, "Some problems in interval estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1954, pp. 175–185.

Fowlie, Meredith, Mar Reguant, and Stephen P Ryan, "Market-based emissions regulation and industry dynamics," *Journal of Political Economy*, 2016, *124* (1), 249–302.

_ , _ , and Stephen P. Ryan, "Measuring Leakage Risk," 2016. Online at https://www.arb.ca.gov/cc/capandtrade/meetings/20160518/ucb-intl-leakage.pdf. Last visited December 2016.

Friedman, Nicole, "Why the Drop in Oil Prices Caught So Many by Surprise," *The Wall Street Journal*, 2014, *October 28.* Online at http://www.wsj.com/articles/why-the-drop-in-oil-prices-caught-so-many-by-surprise-1414526075. Last visited December 2016.

Furman, Jason, "The Decrease in Oil Consumption and Its Impact on the Economy," 2016. Remarks delivered at the Columbia Center for Global Energy Policy on April 28, 2015. Online at https://www.whitehouse.gov/sites/default/files/docs/20150428_oil_consumption_decrease_columbia_center_for_global_energy_policy.pdf.

Gold, Russell, "Drillers Unleash 'Super-Size' Natural Gas Output," *The Wall Street Journal*, 2015, *September 1.* Online at http://www.wsj.com/articles/drillers-get-super-size-natural-gas-output-1441127955?mg=id-wsj. Last visited December 2016.

Gray, Wayne, Joshua Linn, and Richard Morgenstern, "Employment and Output Leakage under California's Cap-and-Trade Program," 2016. Online at https://www.arb.ca.gov/cc/capandtrade/meetings/20160518/rff-domestic-leakage.pdf. Last visited December 2016.

Griliches, Zvi and Jerry A Hausman, "Errors in variables in panel data," *Journal of Econometrics*, 1986, *31* (1), 93–118.

Hamilton, Stephen F., Ethan Ligon, Aric Shafran, and Sofia Villas-Boas, "Production and Emissions Leakage from California's Cap-and-Trade Program in Food Processing Industries: Case Study of Tomato, Sugar, Wet Corn and Cheese Markets," 2016. Online at https://www.arb.ca.gov/cc/capandtrade/meetings/20160518/calpoly-food-process-leakage.pdf. Last visited December 2016.

Hausman, Jerry, "Contingent valuation: from dubious to hopeless," *The Journal of Economic Perspectives*, 2012, *26* (4), 43–56.

Ho, Daniel E, Kosuke Imai, Gary King, and Elizabeth A Stuart, "Matching as

Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 2007, pp. 199–236.

International Energy Agency (IEA)

International Energy Agency (IEA), "Natural Gas Market Review," 2009. Online at https://www.iea.org/publications/freepublications/publication/gasmarket2009.pdf. Last visited December 2016.

Joskow, Paul M., "Natural Gas: From Shortages to Abundance in the United States," *American Economic Review: Papers and Proceedings*, 2013, *103* (3), 1187–1214.

Levy, Armando, 2016. Correspondence to California Air Resources Board Chairman Mary Nichols, June 10, 2016.

Linn, Joshua, "Energy Prices and the Adoption of Energy-Saving Technology," *The Economic Journal*, 2013, *118*, 1986–2012.

– , "Why Do Oil Shocks Matter? The Role of Inter-Industry Linkages in U.S. Manufacturing," *Economic Inquiry*, 2013, *47*, 549–567.

McKinnish, Terra, "Panel data models and transitory fluctuations in the explanatory variable," *Advances in Econometrics*, 2008, *21*, 335–358.

National Regulatory Research Institute (NRRI)

National Regulatory Research Institute (NRRI), "Utility Involvement in Distributed Generation: Regulatory Considerations White Paper," 2015. Online at http://pubs.naruc.org/pub/536EF3DF-2354-D714-510E-B20295D42469. Last visited 2016-11-06.

Olson, Bradley and Erin Ailworth, "Low Crude Prices Catch Up With the U.S. Oil Patch," *The Wall Street Journal*, 2015, *November 20.* Online at http://www.wsj.com/articles/low-crude-prices-catch-up-with-the-u-s-oil-patch-1448066561. Last visited December 2016.

Paul, W. Bernstein, David Montgomery, and Sugandha D. Tuladhar, "Comments on Cap-and-Trade Program Public Workshop on Emissions Leakage Studies," 2016.

Pindyck, Robert S and Julio J Rotemberg, "Dynamic factor demands and the effects of energy price shocks," *The American Economic Review*, 1983, *73* (5), 1066–1079.

Scheffe, Henry, "Multiple testing versus multiple estimation, Improper confidence sets, Estimation of directions and ratios," *The Annals of Mathematical Statistics*, 1970, *41* (1), 1–29.

Wooldridge, Jeffrey M, *Econometric Analysis of Cross Section and Panel Data*, MIT press, 2010.

Young, Cristobal and Katherine Kroeger, "Model Uncertainty and Robustness: A Computational Framework for Multi-Model Analysis," 2015. Online at http://web.stanford.edu/~cy10/public/mrobust/Model_Robustness.pdf. Last visited December 2016.

7 Figures

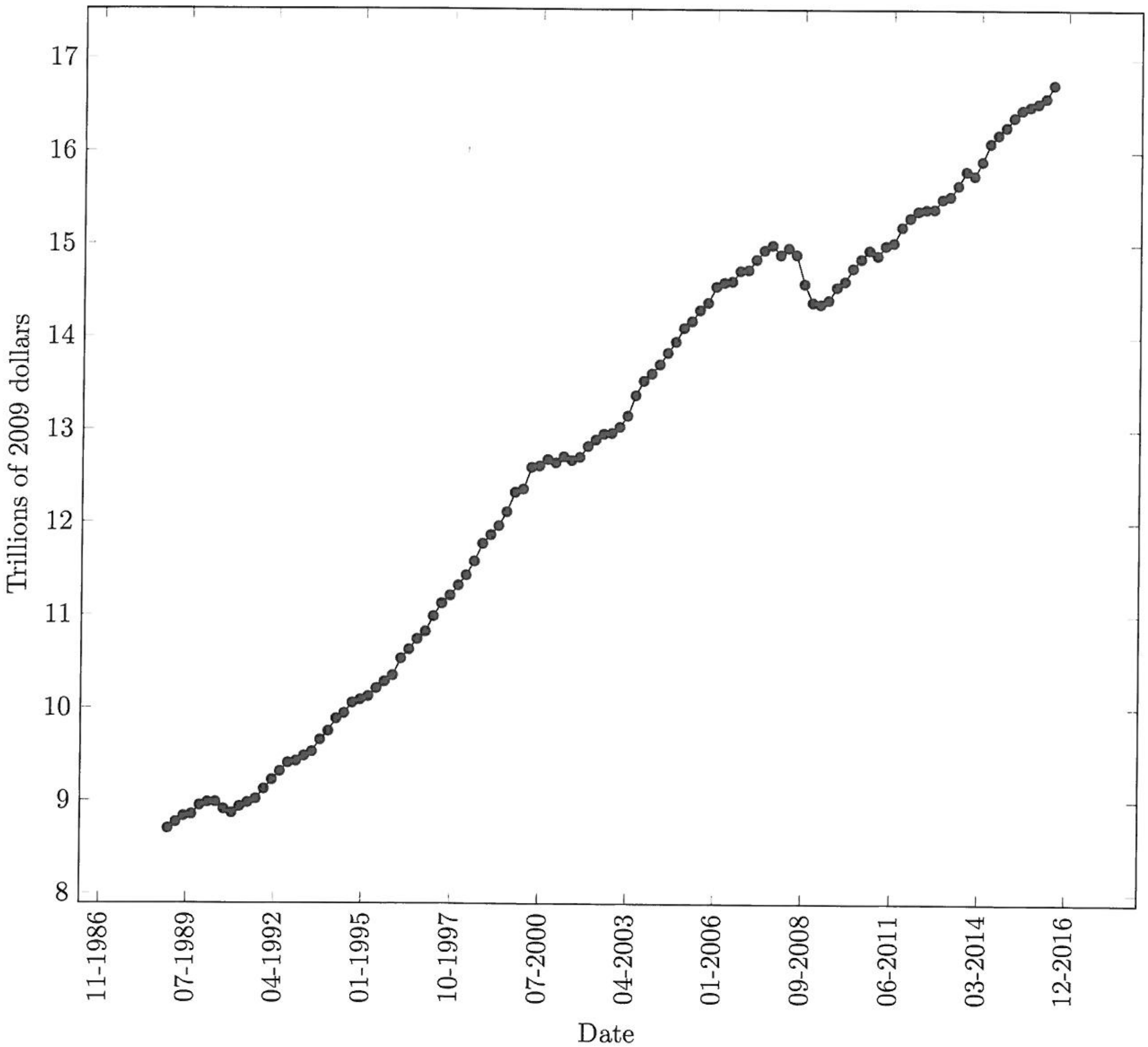
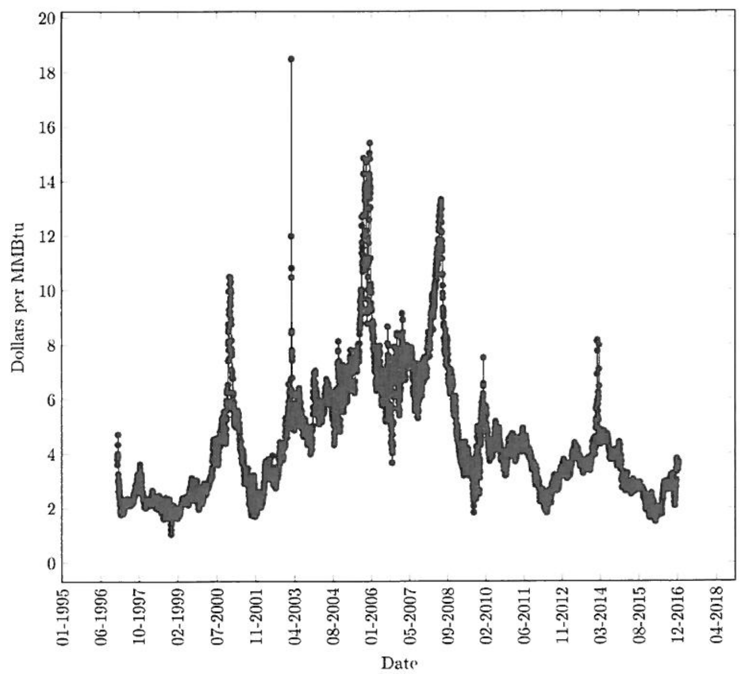
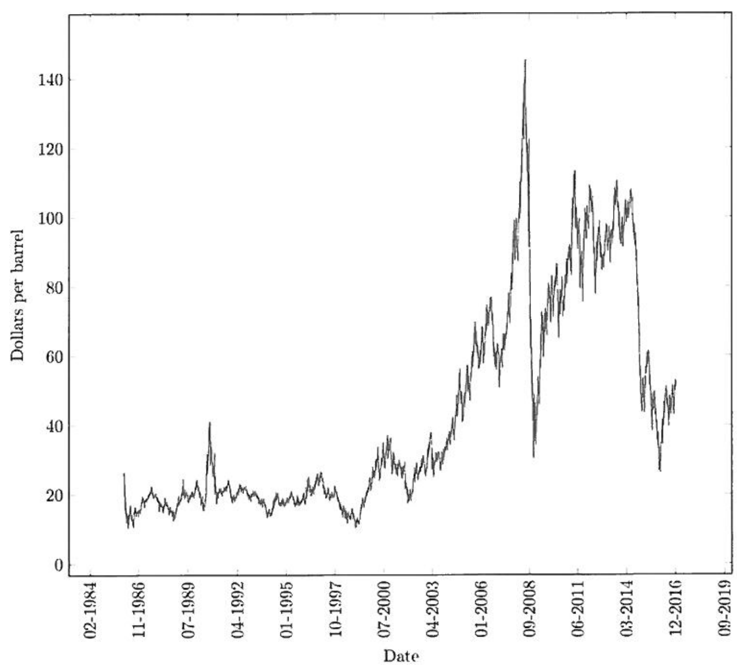


Figure 1: U.S. Real GDP



(a) Henry Hub Natural Gas



(b) WTI Crude Oil

Figure 2: Energy Spot Prices

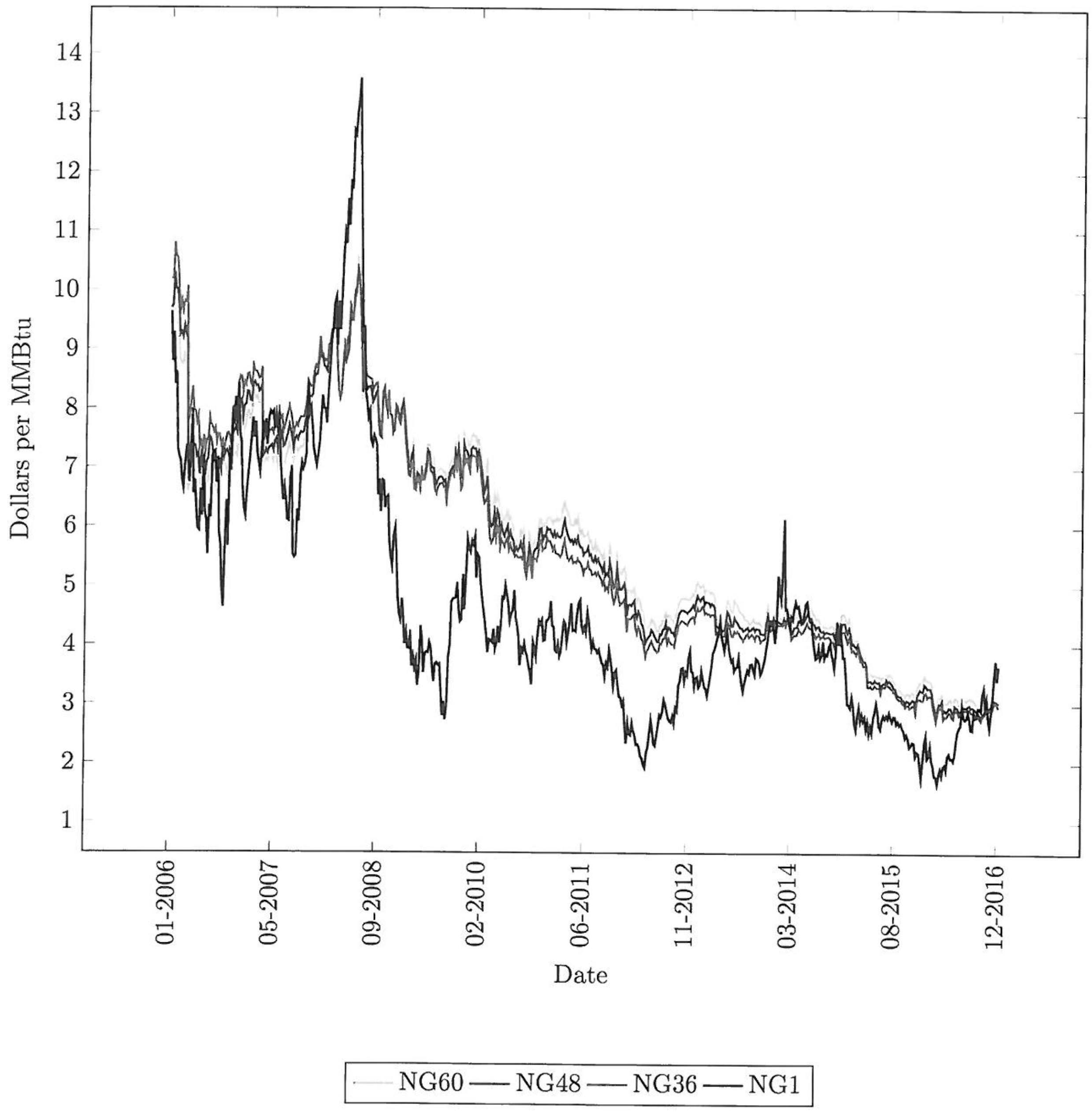


Figure 3: Natural Gas Futures Prices

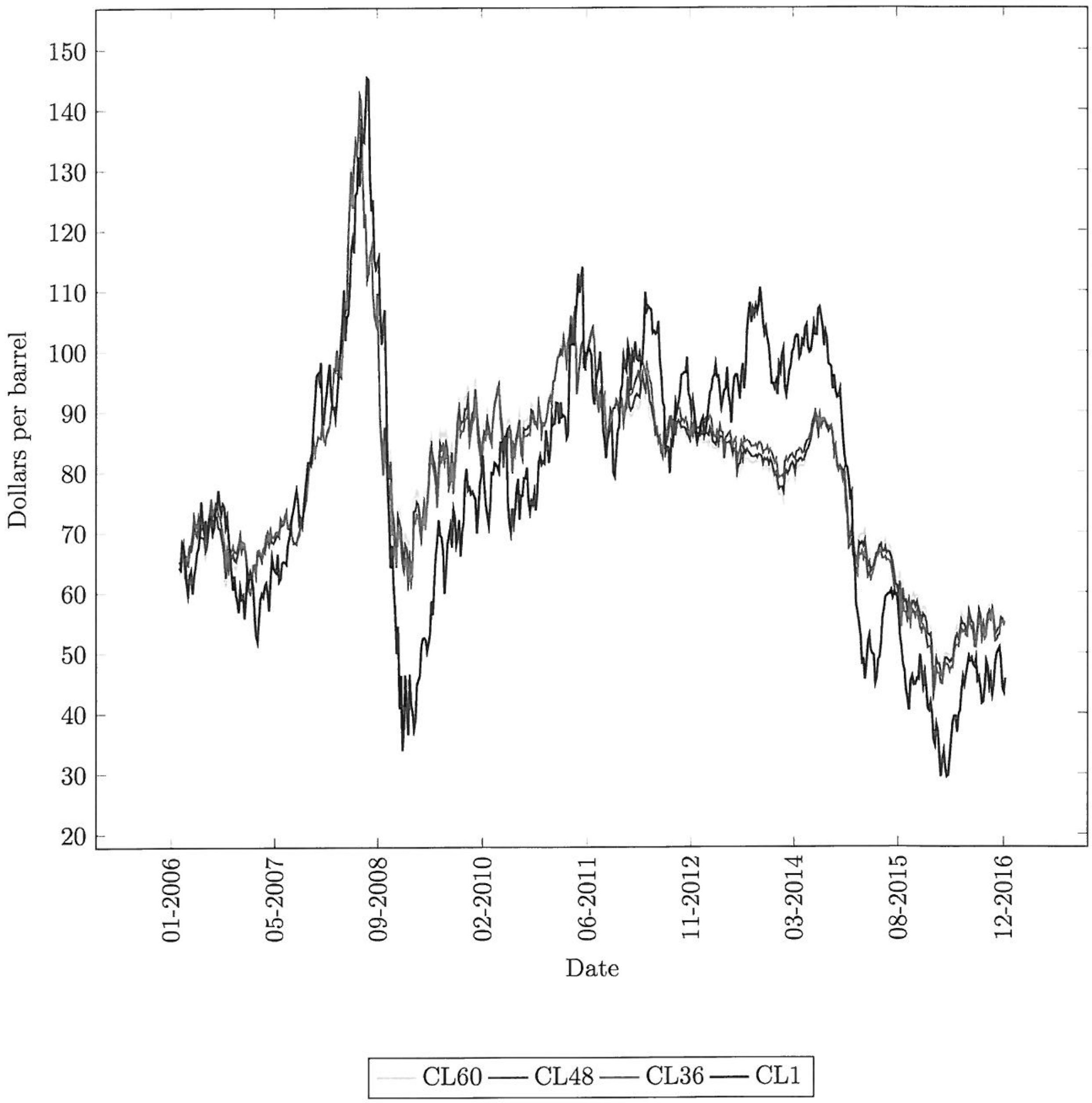
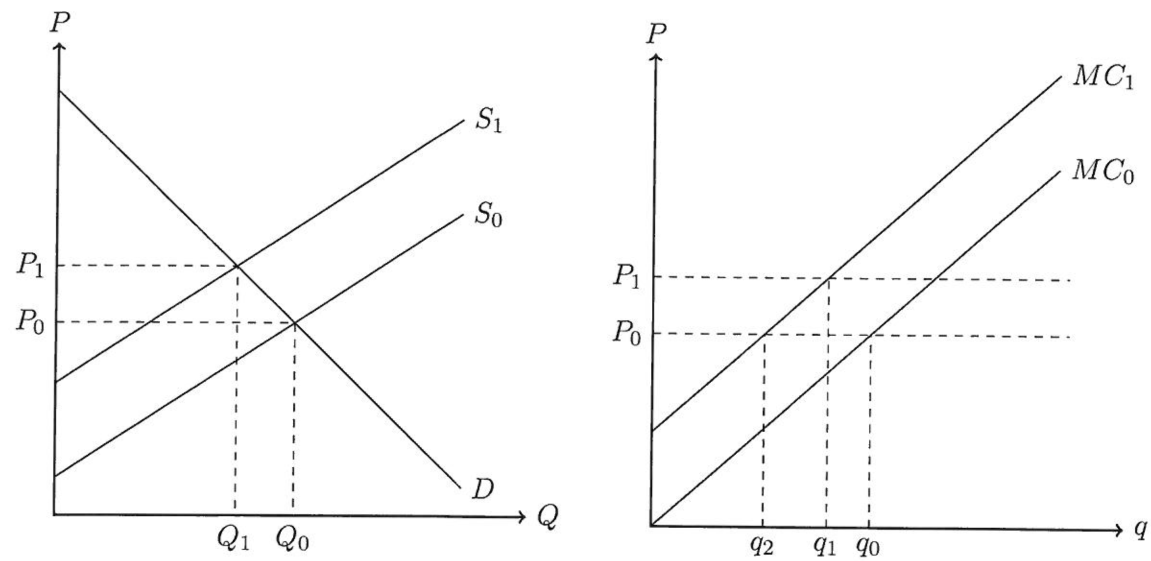
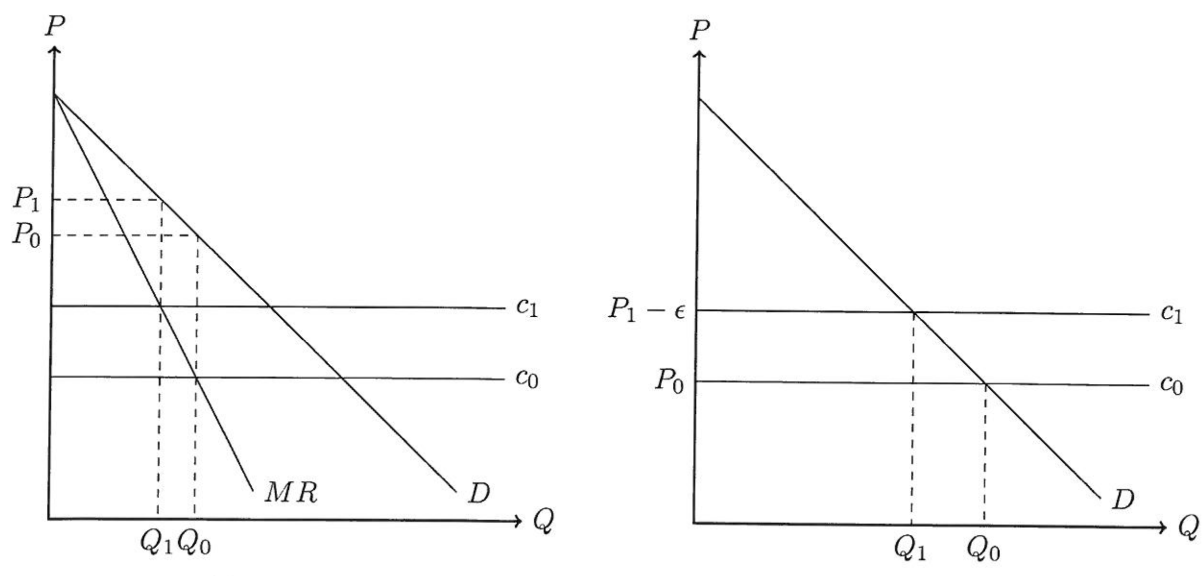


Figure 4: WTI Crude Oil Futures Prices



(a) Market (b) California firm

Figure 5: Output and price changes in response to industry-wide energy cost shock vs. a California-only energy cost shock



(a) Monopoly (b) Betrand duopoly

Figure 6: Monopoly vs. Bertrand duopoly: an identical cost shock induces drastically different output responses