Hotelling Under Pressure*

Soren T. Anderson† Ryan Kellogg‡ Stephen W. Salant§

October 22, 2013

DRAFT: PLEASE DO NOT CITE OR DISTRIBUTE

Abstract

We show that crude oil production from existing wells in Texas does not respond to current or expected future oil prices, contradicting a basic prediction of Hotelling's (1931) canonical model of exhaustible resource extraction. In contrast, the drilling of new wells exhibits a strong price response, as do the rental rates on drilling rigs. To explain these observations, we reformulate Hotelling's model as a drilling problem, in which firms choose when to drill new wells, but flow from existing wells is limited by a capacity constraint that decays exogenously toward zero as reservoir pressure declines. This drilling problem implies a modified Hotelling rule for discounted revenue flows net of drilling costs. Our model rationalizes the empirical findings from Texas and can replicate several other well-known features of the oil industry: local production peaks, backwardated price expectations following positive demand shocks, and expectations that prices will rise faster than the interest rate following large negative demand shocks.

JEL classification numbers: Q3, Q4
Key words: crude oil; supply elasticity; exhaustible resource extraction; Hotelling model

*For helpful comments and suggestions, we are grateful to Jinhua Zhao and conference audiences at AERE, the Energy Institute at Haas Energy Camp, the MSU/UM Energy and Environmental Economics Summer Workshop, and the Occasional California Workshop in Environmental and Resource Economics. Excellent research assistance was provided by Dana Beuschel.

†Michigan State University and NBER. Email: sta@msu.edu. Web: http://www.msu.edu/∼sta.

‡University of Michigan and NBER. Email: kelloggr@umich.edu. Web: http://www.umich.edu/∼kelloggr.

§University of Michigan and Resources for the Future. Email: ssalant@umich.edu. Web: http://www.umich.edu/∼ssalant.

1 Introduction

Ever since Hotelling's (1931) seminal article, economists have modeled the optimal extraction of an exhaustible resource as a cake-eating problem, in which extraction today is traded off directly against the opportunity cost of extracting tomorrow. A common application of the Hotelling logic is to production from an oil reserve. We argue, however, that observed patterns of oil production and prices are not compatible with Hotelling (1931) nor any of its subsequent modifications in the literature. Instead, we show that to replicate structurally the dynamics of oil supply, economists must recognize the physical constraints on oil extraction rates that are imposed by fluid flow dynamics and declines in underground pressure, recasting the Hotelling model as a well-drilling problem (or a keg-tapping problem, if one wishes to maintain an analogy to food and drink) rather than a cake-eating problem.

We begin by showing empirically, using data from Texas over 1990–2007, that oil production from drilled wells declines asymptotically toward zero and is not affected by shocks to spot and expected future oil prices. We also show empirically that this behavior, which is inconsistent with extraction models in the literature, is not simply driven by institutional factors, such as common-pool problems or oil lease contract provisions. Instead, the observed decline in production is rationalized by the cost structure of the oil industry and by the flow constraints imposed by underground pressure dynamics. When a well is first drilled, the pressure in the underground oil reservoir is high. Thus, production may initially be rapid, since the rate of fluid flow is roughly proportional to the pressure difference between the reservoir and the surface. Over time, however, the reservoir pressure abates as less and less of the original oil remains, and the well's flow must gradually decay toward zero. Thus, while extractors can choose when to drill their wells, the maximum rate of production from these wells is geologically constrained to decline with the total quantity extracted. Because the marginal cost of producing oil from a drilled well is very low relative to oil prices and drilling costs, the flow constraint typically binds in equilibrium, yielding the asymptotic decline in production that we observe in the data.

While production from drilled wells is insensitive to oil prices, our Texas data also show that the drilling of new oil wells and the rental price of drilling rigs both respond strongly to oil price shocks. These findings imply an upward-sloping supply curve for drilling rigs and crews to operate them. These empirical results motivate us to recast Hotelling's (1931) canonical model as a drilling problem rather than a production problem. Consistent with our empirical findings, extractors in our model choose when to drill their wells, but the flow from these and other existing wells decays asymptotically toward zero at an exogenous rate. Thus, there are two state variables in our model: (1) the flow of oil from existing wells, and (2) the number of wells remaining to be drilled.

We initially consider the case of a small number of large wells, showing that the planner would choose a saw-tooth marginal utility path that cannot be decentralized as a competitive (price-taking) equilibrium. Since there are millions of operating wells in the world, we focus on the more realistic case of many infinitesimally small wells. In this case, the planner's problem can be decentralized as a competitive equilibrium. We show that a modified Hotelling rule holds in our model: whenever drilling occurs, the discounted revenue stream that flows from each well, less the marginal cost of drilling it, rises at the rate of interest.

We show that the traditional Hotelling path can, via judicious control of the drilling rate, still emerge as optimal—but only under the assumptions that marginal drilling costs are constant and that the marginal utility of consuming oil is unbounded as consumption declines to zero. The first assumption is contradicted by our data on rental rates for drilling rigs, and the second is dubious given the existence of viable alternative fuel sources.

When we relax these two assumptions, our model easily and naturally replicates key qualitative features of the crude oil extraction industry. We first consider the behavior of private oil developers in a local region facing an exogenous price of oil and an upward-sloping local supply of drilling rigs for rent. This simple model generates the classic "peak oil" production profile observed at every level of aggregation, from individual oil fields to large oil-producing regions (Hamilton 2013). Production initially rises as developers in a region

rapidly drill wells but then inevitably declines as drilling slows and the flow from existing wells decays. Consistent with our data from Texas, an unanticipated, sudden increase in the oil price leads to an immediate increase in drilling activity and rig rental rates, shifting the peak in production earlier in time and making it more pronounced. Thus, a global price increase could slow or even temporarily reverse a downward trend in production, as we have seen throughout oil-producing regions of the United States in recent years (Hamilton 2013).

We next consider the case of endogenous oil prices and a general supply curve for the rental of drilling rigs that is upward sloping. We begin by considering an analytically tractable case in which there is an unlimited number of wells to be drilled—an approximation to a world in which the stock of oil is vast relative to current demand. In this case, our model reduces to a standard macroeconomic model of optimal investment in the presence of convex adjustment costs, with the decaying flow from existing wells playing the role of an industry's depreciating capital stock and the drilling of new wells playing the role of investment. Positive oil demand shocks lead to an immediate increase in oil prices, drilling, and rig rental prices, with oil prices then falling as production builds gradually to a new steady state. These results are reversed for negative demand shocks, which can—if large enough—cause oil prices to rise faster than the rate of interest temporarily following the initial drop in prices, even though production and drilling need not fall all the way to zero. Thus, our model generates the same qualitative responses to demand shocks that we observe in real-world production data and futures markets.

Lastly, we consider the case of endogenous oil prices in the presence of scarcity. For tractability, we assume that rigs can be rented at a constant price up to a drilling capacity constraint. This model also generates a peak in oil production, with production ramping up and prices falling as well owners race to drill at the constraint. Though well revenues are falling initially, scarcity rents earned by rig owners are also falling, leaving well owners indifferent as to when they drill. Eventually, drilling drops below the constraint, and the price of oil less the per-barrel marginal cost of drilling rises at the rate of interest until

the last well is drilled, at which point the oil price rises exogenously to the upper bound on prices. Numerical simulations demonstrate that qualitatively similar results emerge for more general, upward-sloping supply functions for drilling rig rentals.

Our paper is not the first to modify or extend Hotelling's original model.[1] However, our study departs from previous work in several important ways. First, we present an extensive set of empirical results to motivate and justify every assumption in our model. Our key empirical finding—that oil production from drilled wells is unresponsive to price shocks—has not, to the best of our knowledge, been documented in prior work. Second, we fundamentally recast the oil extraction problem from one in which the rate of production is the central choice variable to one in which extractors choose the rate at which new wells are drilled. This change in choice variable arises naturally from our empirical work and the binding production constraint that is implied by the declining pressure of the oil reservoir. Third, we show that our reformulated extraction problem yields production and price dynamics that can match observed production data and price expectations from futures markets: local oil-producing fields and regions exhibit production peaks, while expected future oil prices can be backwardated after positive demand shocks and can rise faster than the interest rate following negative demand shocks.

Within the Hotelling literature, our paper is most closely related to papers featuring models in which firms face convex costs—often motivated as "set-up" or "exploration" or "site-development" costs—to expand reserves or production capacity (Pindyck 1978; Switzer and Salant 1986; Livernois and Uhler 1987; Holland 2008). Such models also generate an initial period of rising production and falling prices as capacity builds. Prices eventually

[1]The economics literature following Hotelling (1931) models, among other things: extraction costs that increase with the rate of production from a reserve or over time as a reserve is depleted; heterogeneity in extraction costs across different reserves; cost-reducing technological change; demand growth; extraction by firms with market power; exploration and development of new reserves; and uncertainty of various kinds (e.g., in demand or resource discovery). For a review of the theoretical literature through the 1970s, see Devarajan and Fisher (1981). For a more recent review, see Gaudet (2007). For an accessible theoretical primer, see Salant (1995). In addition, a multiple demand curve model has been developed to take account of users in different countries who must pay different transport costs or taxes for the same resource or must pay to convert one resource before using it as a substitute for another type of resource. For a synthesis of this recent literature on "multiple demand curves" see Gaudet and Salant (2014).

rise as production inevitably declines, but the capacity constraint—if any—remains set to its previous maximum and therefore does not bind. Thus, extractors in these other models, operating below their constraint, can arbitrage price shocks by accelerating or deferring production freely. In contrast, the production constraint in our model shifts relentlessly toward zero whenever production occurs, while cumulative drilling to rebuild capacity is bounded by the finite number of wells.[2] As a result, the production constraint always binds in our model, even during periods in which production is falling, and drilling remains the only mechanism by which extractors are actively able to control production. Thus, in contrast to existing models, our model can explain why production from drilled wells declines steadily over time yet simultaneously does not respond to price shocks, and why futures markets sometimes indicate that prices are expected to rise faster than the rate of interest. In short, while the long-run price paths emerging from these models are qualitatively similar to ours, the underlying mechanism of production is different, and the models make very different predictions for short-run dynamics in response to unanticipated shocks.

We also contribute to an extensive empirical literature testing whether the Hotelling rule holds in practice, generally finding that it does not.[3] Of course, there are countless reasons why this rule may fail to hold in practice, given the original model's many simplifying assumptions and the difficulty of specifying the parameters of a more realistic model. Thus, rather than test the precise numerical implications of a model that we know is unrealistic, we simply ask whether crude oil production responds *at all* to current and expected future

[2]To the best of our knowledge, only two other papers in the economics literature—Thompson (2001) and a recent, unpublished working paper, Mason and van't Veld (2013)—are premised on the idea that the production constraint falls over time. The former is primarily concerned with valuing oil reserves, while the latter focuses on the value of enhanced oil recovery. Our paper differs in that: (1) we show empirically that the flow constraint has historically been binding and discuss why this should be the case in equilibrium; (2) we develop a new resource extraction model that incorporates this constraint and explicitly focuses on the drilling of wells, allowing for convex drilling costs as motivated by our empirical results; and (3) we extensively explore the short-run and long-run dynamics of drilling, production, and prices implied by our model.

[3]Smith (1979), Slade (1982), and Berck and Roberts (1996) find limited evidence for an upward trend in exhaustible resource prices, but tests based on price alone are not correctly specified unless extraction costs are negligible. Structural econometric papers estimating in-situ values, including Miller and Upton (1985), Halvorsen and Smith (1984), Black and LaFrance (1998), and Thompson (2001), find mixed results. See Slade and Thille (2009) for a recent review.

prices in the way that such a model would predict. As we show, new drilling activity increases with spot prices, but production from existing wells is almost entirely price inelastic. These results motivate our radical rethinking of the canonical model to focus on drilling as the central mechanism of production.

Finally, we contribute to a broader empirical literature estimating the response of crude oil production to prices. This work has generally found that production in aggregate is price inelastic, at least in the short run (Griffin 1985; Hogan 1989; Jones 1990; Dahl and Yucel 1991; Ramcharran 2002), though (Rao 2010) finds evidence using well-level data that firms can shift production across wells in response to well-specific taxes. The empirical response of crude oil supply to prices has important implications for the macroeconomic effects of oil supply and demand shocks, since inelastic supply and demand lead to volatile prices. See Hamilton (2009) and Kilian (2009) for insightful discussions of these issues. Our empirical findings are consistent with a highly inelastic short-run supply of Texas crude oil, with somewhat more elastic supply over the medium run as drilling activity responds.[4] Our theoretical results suggest that this medium-run response comes at the expense of lower production in the more distant future—as it must, for a finite resource.

The remainder of the paper proceeds as follows. Section 2 presents our empirical evidence on crude oil production and drilling in Texas. Section 3 recasts Hotelling's model as a drilling problem, initially considering the case of a small number of large wells before turning to the more realistic case of a large number of infinitesimally small wells. Section 4 shows that it is possible to replicate the Hotelling path exactly, but only under implausible assumptions, including constant marginal drilling costs. Section 5 relaxes this assumption and considers three cases: drilling in a local region that faces an exogenous price path of oil, a closed model that endogenizes the price of oil in the absence of resource scarcity, and finally a closed model with scarcity. Section 6 then concludes.

[4]Kline (2008) shows that wages and employment in the oil and gas field services industry also increase with crude oil prices—though these labor market effects emerge more slowly.

2 Empirical evidence from Texas

In this section we study how oil production and drilling in Texas respond to incentives generated by changes in current and expected future oil prices, as revealed in futures markets. We show that production exhibits nearly zero response to current and expected future prices, whereas drilling activity—along with the cost of renting drilling rigs—responds strongly. We then discuss how these results derive from the fundamental cost structure and technology of the crude oil extraction industry, and we rule out several alternative explanations.

2.1 Data sources and data cleaning

Our crude oil drilling and production data come from the Texas Railroad Commission (TRRC), covering the period 1990–2007. The drilling data come from the TRRC's "Drilling Permit Master" dataset, which provides the date, county, and lease name for every well drilled in Texas. A lease is a plot of land upon which an oil production company has obtained from the (usually private) mineral rights owner the right to drill for and produce oil and gas. Over 1990–2007, a total of 157,271 new wells were drilled, along with 42,893 "re-entries" of existing wells.[5]

The production data come from the TRRC's "Oil and Gas Annuals" dataset, which records monthly crude oil production at the lease-level.[6] The TRRC production data are at the lease-level rather than the well-level because individual wells are not flow-metered.[7]

[5] A re-entry occurs when a rig is used to deepen the well, drill a "sidetrack" well off of the existing well bore, or attempt to stimulate production through perforating or fracturing the oil reservoir—possibly at a different depth than that from which the well was previously producing. These interventions are all similar to (though somewhat cheaper than) drilling a new well in that they require a substantial up-front capital investment and allow access to a previously untapped section of oil-bearing rock.

[6] Due to false zeros in the raw dataset provided by the TRRC for some leases in 1996 and December 2004-2007, we augmented these data with production information scraped from the TRRC's online production query tool. We verified that the online data matched the raw dataset for leases and months not affected by the data error.

[7] Direct production from a well is a mix of oil, gas, and often water. These products must first be separated before metering of oil flow can take place; this separation typically takes place at a single facility that serves all of the wells on the lease. Oil flows from the separation facility into storage tanks, from which production is metered when it is delivered to either a pipeline or truck for sale. Although firms often assess well-level productivity monthly by diverting each well's flow into a small "test separator," these data are not available from TRRC nor would they be particularly reliable for accurately measuring a well's flow over the course of

Thus, we generally cannot observe well-level production, though for some analyses we will take advantage of a set of leases that have a single well.

Our analysis of the production data focuses on whether firms respond to oil price shocks month-to-month by adjusting the flow rates of their existing wells (by speeding up or slowing down the pumping unit),[8] or shutting in (turning off) or restarting their wells. To distinguish these actions from investments in new production capital, such as drilling a new well, we discard from our production data leases in which any rig work took place.[9] In the remaining data, there exist 16,148 leases for which production data are not missing for any month from 1990–2007 and production is non-zero for at least one month, so that our dataset consists of 3,487,968 lease-months of production. The typical oil lease in Texas has a fairly low rate of production, reflecting the fact that oil fields in Texas are mature and have been heavily produced in the past. The average daily lease production in the data is 3.6 barrels of oil per day (bbl/d), with a standard deviation of 18.2 bbl/d. A total of 1,070,632 (31%) of the observed lease-months have zero production, and the maximum observed production is 9,510 bbl/d.

Our oil price data come from the New York Mercantile Exchange (NYMEX) and measure prices for West Texas Intermediate (WTI) crude oil delivered in Cushing, Oklahoma—the most common benchmark for crude oil prices in North America—from 1990–2007. We use the Bureau of Labor Statistics's All Urban, All Goods Consumer Price Index (CPI) to convert all prices to December 2007 dollars. We use the front-month (upcoming month) futures price as our measure of the spot price of crude oil and use prices for longer-term futures contracts to measure firms' price expectations.

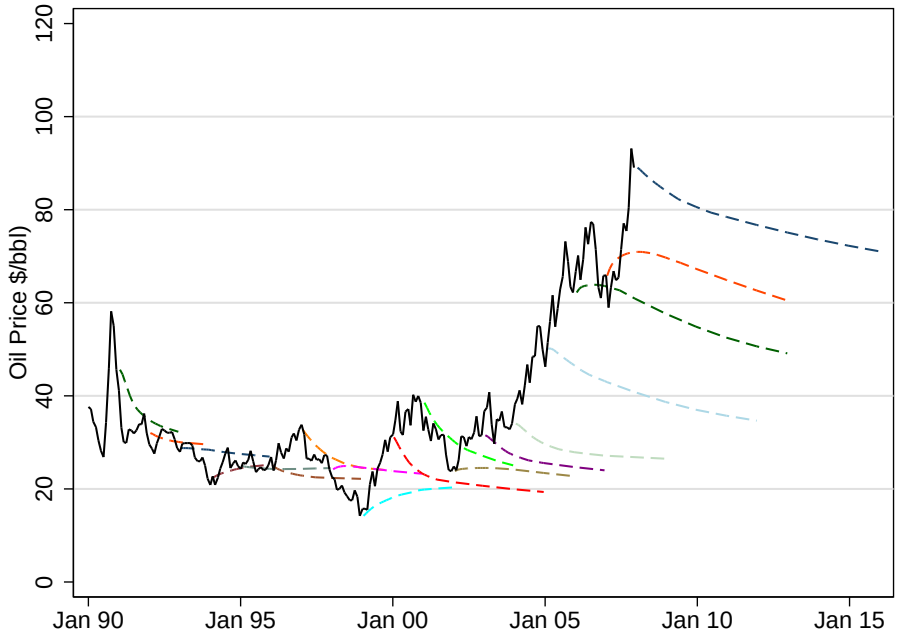
---

an entire month.

[8]The vast majority of the wells in the dataset are pumped and do not flow naturally. The average lease-month in the data has 2.02 pumped wells and 0.06 naturally flowing wells.

[9]Discarding these leases requires matching the drilling dataset to the production dataset. Accomplishing this match at the lease level is difficult because the datasets must be matched on lease name, which is not consistent across the two datasets. Rather than risk having our production dataset contaminated by leases in which rig activity occurred, we instead conservatively identify all county-firm pairs in which rig work took place and then discard *all* leases corresponding to those county-firm pairs (counties and firms are identified with numeric codes in both datasets and therefore do not suffer match problems).

Figure 1: Crude oil spot prices and futures curves



Note: This figure shows crude oil spot and futures prices, all in real $2007. The solid black line represents the spot price of oil. The dashed colored lines represent the futures curves as of December in each year. See text for details.

Figure 1 shows the time series of crude oil spot prices (solid black line) as well as the futures curves as of December in each year (dashed colored lines).[10] For example, the left-most dashed line shows prices in December 1990 for futures contracts with delivery dates from January 1991 through December 1992. As is clear from the figure, the futures market for crude oil is usually backwardated, indicating that traders typically expect the future price to be lower than the current spot price. However, the figure also shows several periods of contango, particulary during the 1998–1999 low price period, in which the futures curve is upward sloping.

[10]We have converted all of the price data so that the slopes of the futures curves in figure 1 and the expected rates of price change used in our analysis reflect real rather than nominal changes. To convert the futures curves from nominal to real, we adjust for both the trade date's CPI and for expected annualized inflation of 2.5% between the trade date and delivery date. The average annual inflation rate from January 1990 to December 2007 is 2.50%, and inflation varies little over the sample. Thus, for example, we convert the nominal prices for futures contracts traded in December 1990 to real price expectations by multiplying by the December 2007 CPI, dividing by the December 1990 CPI, and then dividing each contract price by $1.025^{t/12}$, where t is the number of months between the trade date and the delivery date.

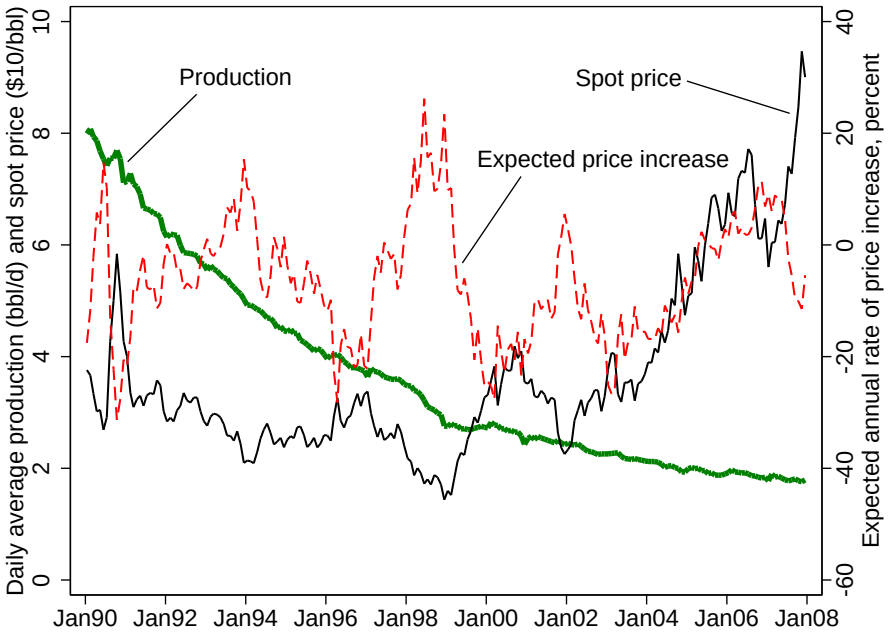
The use of futures prices to measure price expectations in not without controversy. Alquist and Kilian (2010), for example, shows that futures markets do not out-perform a simple no-change forecast in out-of-sample forecasting over 1991-2007. We use futures markets here for several reasons. First, NYMEX futures are liquidly traded at the horizons we consider here, and with many deep-pocketed, risk-neutral traders, the futures price should equal the expected future spot price. Second, a majority of oil producers in Texas claim to use futures prices in making their own price projections (Society of Petroleum Evaluation Engineers 1995). Third, Kellogg (Forthcoming) shows that the drilling activity of Texas oil production firms is more consistent with price expectations based on the futures market than on a no-change forecast. Finally, as we discuss below, we find that firms' on-lease oil stockpiles increase when futures prices are high relative to spot prices, as one would expect if firms' expectations aligned with the market.

Finally, we have also obtained information on rental prices ("dayrates") for drilling rigs from RigData, a firm that publishes reports on the U.S. onshore oil and gas industry. As discussed in Kellogg (2011), the oil production companies that make drilling and production decisions do not drill their own wells but rather contract drilling out to independent service companies that own rigs. Paying to rent a rig and its crew typically constitutes the largest line-item in the overall cost of a well. The data provided by RigData are quarterly, covering Q4 1990 through Q4 2007, and are broken out by region and rig depth rating. We use dayrates for rigs with depth ratings between 6,000 and 9,999 feet (the average well depth in our drilling data is 7,425 feet) for the Gulf Coast / South Texas region. Observed dayrates range from \$6,315 to \$15,327 per day, with an average of \$8,008 (all real December 2007 dollars).

2.2 Production from existing wells does not respond to prices

Our main empirical results focus on the production data for leases on which there was no rig activity from 1990–2007, so that all production in the data come from pre-existing

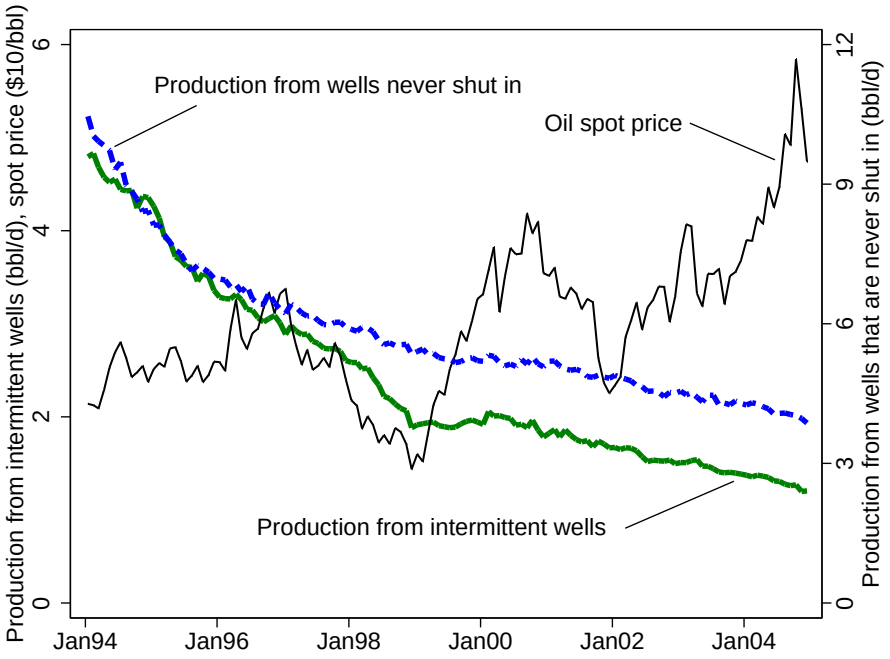
Figure 2: Crude oil prices and production from existing wells



Note: This figure presents crude oil spot prices and the expected percent change in prices over one year, as well as daily average production from leases on which there was no rig activity (so that all production comes from pre-existing wells). All prices are real $2007, and expected price changes are net of inflation. See text for details.

wells. Figure 2 presents daily average production (in bbl/d) for these leases, along with crude oil spot prices and the expected percent change in spot prices over one year. Average production is dominated by a long-run downward trend, with little apparent response either to the spot price of oil or to the expected percent change in prices over one year. This result contrasts sharply with predictions from Hotelling models that are standard in the literature, which would predict a complete shutdown of production during periods, such as 1998–1999, when prices are expected to rise more quickly than the rate of interest. Moreover, under a commonly used assumption of increasing marginal extraction costs, Hotelling models will predict that production should increase with the spot price and decrease with the expected future change in price. None of these predictions appears strongly in the time series of production from pre-existing wells. In addition to the graphical evidence presented here, we

Figure 3: Intermittent wells versus wells never shut in



Note: This figure shows production from wells that shut in at least once during 1994–2004 versus production from wells that never shut in during 1994–2004. Production data come from leases that had no more than one productive well and never experienced a rig intervention over 1994–2004. Oil prices are real $2007. See text for details.

have found that the lack of response to price incentives holds broadly both for high-volume wells and for naturally flowing wells, indicating that our main empirical result is not specific to low-volume pumped wells that are most common in Texas. We have also confirmed the overall lack of response to price incentives using a formal regression analysis.[11]

Figure 2 suggests that production did deviate slightly from the long-run trend during the 1998–1999 period during which the spot price fell below $20/bbl and the expected rate of price change exceeded 10% (and sometimes 20%). In particular, it appears that production accelerated its decline rate in 1998 while prices were falling, leveled off in 1999 while prices were rising, and then resumed its usual decline in 2000.

To assess whether this deviation is real and what mechanism lay behind it, we study

[11] Ancillary results not presented in the paper, both here and below, are available upon request.

whether it arose from wells being shut in or changes in production from active wells. We first isolate the sample to leases that had no more than one flowing well over 1994–2004, so that observed lease-level production during this time can be interpreted as well-level production. We then split this sample into two groups: wells that are never shut in over 1994–2004 and "intermittent" wells that are shut in at least once. Figure 3 plots the time series of production from these two samples. This figure makes clear that the 1998 deviation from trend was driven entirely by marginal wells that sometimes have zero production. For wells that always produce, there is no adjustment on the intensive margin. It appears that when prices fell in 1998, an unusually large number of wells were shut in, temporarily accelerating the decline. Then, when prices recovered during 1999, many of these wells were returned to production, temporarily slowing the decline. Apart from these deviations, a large response of production to price signals does not appear anywhere in the data.

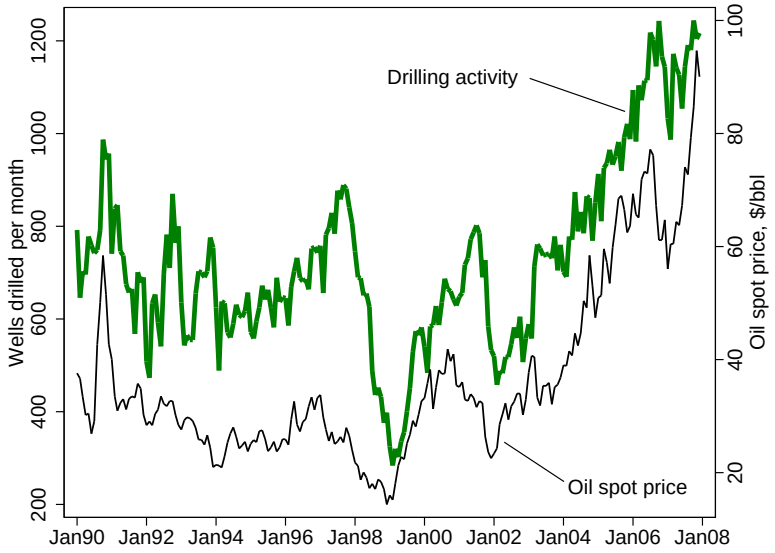
2.3 Rig activity does respond to price incentives

These no-response results based on existing wells stand in stark contrast to new drilling activity in Texas. Figure 4(a) shows the total number of new wells drilled across all leases in our dataset, along with the spot price for crude oil. The figure shows a pronounced monthly correlation between oil prices and new drilling activity. In addition to the graphical evidence presented here, we have found using regression analysis that, for every $1 increase in crude oil prices, the monthly rate at which new wells are drilled increases by a strongly statistically significant 1%–1.5%, while the relationship between drilling and the expected percent change in prices is negative but insignificant. We have also found that the use of rigs to re-enter old wells correlates with oil prices, though not as strongly as the drilling of new wells.

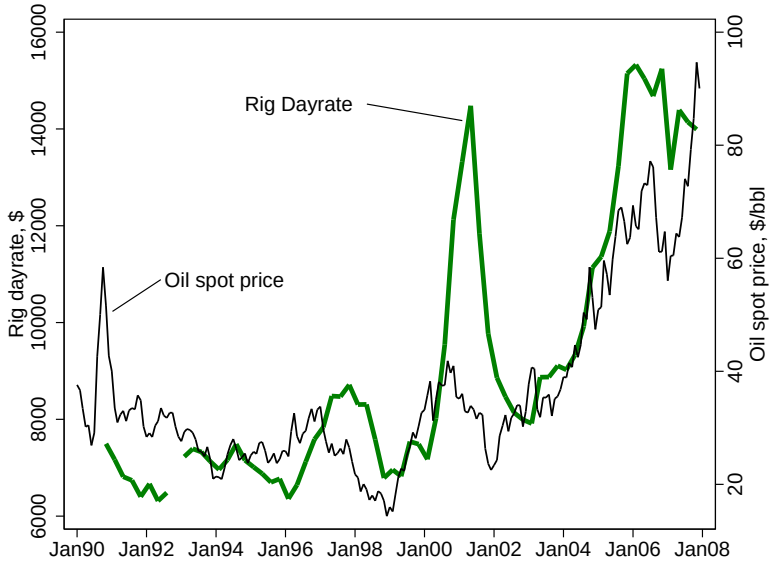
When oil production companies drill more wells in response to an increase in oil prices, more rigs (and crews) must be put into service in order to drill them. Figure 4(b) shows that these fluctuations in rig demand are reflected in a positive covariance between rig dayrates and oil prices. In addition to the graphical evidence presented here, we have found using

Figure 4: Rig activity versus crude oil spot prices

(a) Drilling of new wells



(b) Rig dayrates



Note: Panel (a) shows the total number of new wells drilled across all leases in our dataset. Panel (b) shows dayrates for the Gulf Coast / South Texas region, for rigs with depth ratings between 6,000 and 9,999 feet. The dayrate data are quarterly rather than monthly. Data are available beginning in Q4 1990, and data for Q4 1992 are missing. See text for details.

regression analysis that this relationship is statistically significant and has a magnitude such that a $1 increase in the crude oil spot price is associated with a roughly 2% increase in the average rig dayrate. Thus, as the industry collectively wishes to drill more wells within a given time frame, the marginal cost of drilling those wells increases.

2.4 Industry cost structure explains these price responses

The analysis above documents two empirical facts about oil production and drilling in Texas from 1990–2007. First, production from drilled wells is almost completely unresponsive to changes in spot or expected future oil prices, with an exception being an increased rate of shut-ins during the 1998 oil price crash. Second, drilling of new wells responds strongly to oil price changes, and rig dayrates respond commensurately. Here, we argue that these empirical results reflect an industry cost structure with the following characteristics:

1. The maximum rate of production from a well is physically constrained, and this constraint declines asymptotically toward zero as a function of cumulative production; this function is known as a well's production decline curve.

2. The marginal cost of production below a given well's capacity constraint is very small.

3. The fixed costs of operating a producing well are non-trivial. There may also be costs for restarting a shut-in well, but they are often small enough to be overcome.

4. Drilling rigs and crews are a relatively fixed resource, at least in the short run, leading to an upward-sloping supply curve of drilling rigs for rent.

The capacity constraint and low marginal production cost relate to the observation that production from existing wells does not respond to oil price shocks on the intensive margin. Because oil production firms in Texas are price-takers,[12] production will increase with oil

[12]The market for crude oil is global, and Texas as a whole (let alone a single firm) constitutes only 1.3% of world oil production (Texas and world oil production data for 2007 from the U.S. Energy Information Administration); thus, the exercise of market power by Texas oil producers is implausible.

prices only if the marginal cost of producing is upward-sloping and of similar magnitude to price. Because we do not observe such a response in the data, it must be that the oil price intersects marginal cost at a vertical, capacity-constrained section of the curve. While the marginal cost of production below the capacity constraint is not necessarily zero, it must be well below the range of oil prices observed in the data.

The existence of a capacity constraint for well-level production is consistent not only with the data presented above, but also with standard petroleum geology and engineering.[13] As noted recently in the economics literature by Mason and van't Veld (2013), the flow of fluid through reservoir rock and up the well bore is governed by Darcy's Law (Darcy 1856), which stipulates that the rate of flow is proportional to the pressure differential between the reservoir and the well.[14] In the simplest model of reservoir flow, the reservoir pressure is proportional to the volume of fluid in the reservoir. In this case, the maximum flow rate is proportional to the remaining reserves, which results in an exponential production decline curve. More complex cases, which might involve the presence of gas, water, or fractures in the reservoir, may yield a more general hyperbolic decline. Regardless, the physical laws governing fluid flow place a limit on the rate at which oil can be extracted from a reservoir, and this limit declines with cumulative production. In fact, this issue is sufficiently important that entire sub-fields of petroleum engineering—reservoir engineering and decline curve analysis—are devoted to understanding it.

The existence of a capacity constraint also explains why most producers do not shut in their wells during periods of severe contango, such as 1998, when oil prices are expected to rise more quickly than the rate of interest. A standard Hotelling analysis would suggest that producers should shut in at such times and then recover the lost production once price expectations are backwardated or at least in less severe contango. In the presence of a

[13]We give the geologic and engineering basis for well-level capacity constraints only a brief treatment here. For a fuller discussion of fluid flow and production decline curves, Hyne (2001) is an excellent source that does not require a geology or engineering background.

[14]Installing a pump on a well effectively eliminates the need for the oil to overcome gravity as it rises up the well. Darcy's Law still binds on the flow of oil through the reservoir into the bottom of the well.

capacity constraint, however, when a well is restarted after a shut-in, it cannot instantly produce all of the oil that would have been produced were it not for the shut-in. Instead, production returns to its location on the decline curve from the time before the shut-in, so that the shut-in effectively pushes the entire production profile back in time. Thus, for shutting-in a well to be optimal, prices must be expected to rise more quickly than the rate of interest for a long period of time, spanning most of a well's remaining life. Such long-lived contango does not occur in the futures data, as can be seen in figure 1.[15]

Per figure 3, some relatively low-volume wells were shut in during 1998. These shut-ins are consistent with the existence of fixed production costs, which intuitively arise from the need to monitor and maintain surface facilities such as pumps, flowlines, and separators so long as production is nonzero. When the oil price fell in 1998, production from these wells was no longer sufficient to cover their fixed costs, explaining the decision to shut in. When oil prices subsequently recovered, many of these wells restarted, suggesting that start-up costs are relatively minor.

In appendix A, we carefully consider and rule out alternative explanations for the lack of response of oil production to oil prices. We show that the overall lack of price response cannot be explained by (1) well-specific production quotas (because production quotas are not binding); (2) by leasing agreements that require non-zero production (because multiple-well leases show the same results); (3) by races-to-oil induced by open-access externalities within oil fields (because fields controlled by a single operator show the same results); or (4) by producer myopia or price expectations that are not aligned with the futures market (because producers respond to high futures prices by stockpiling oil above ground).

[15]For instance, in December 1998, the price for delivery in December 1999 is 21.4% greater than the front month price, indicating severe short-run contango. However, the price for delivery in December 2000 is only 10.3% greater than that for December 1999, and the price for delivery in December 2001 is only 4.2% greater than that for December 2000. Thus, shutting in a well in December 1998, even for a brief period, causes production to be delayed over time periods when prices are not expected to rise more quickly than the rate of interest. In principle, deferring production can still be optimal if, as in this example, the price is expected to increase sharply and then remain persistently high. Our calculations suggest, however, that such a price increase would need to be several times greater than even this extreme episode.

3 Recasting Hotelling as a drilling problem

In this section, we develop our theory of optimal drilling. We begin by restating the standard Hotelling result for the planner's problem Assume that oil flow $F(t)$ at time t generates instantaneous utility flow of $U(F(t))$, with $U(0) = 0$, $U'(\cdot) > 0$, and $U''(\cdot) < 0$. Assume a social planner discounts utility continuously at exogenous rate r and, if wealth maximizing agents are involved, they also discount profit flows at rate r. Assume that S units of oil can be costlessly extracted at rate $F(t) \geq 0$, which is under the complete control of the oil extractor. To maximize the discounted utility, the planner chooses $F(t)$ so that marginal utility grows exponentially at the rate of interest whenever oil flow is strictly positive:

$$F(t) \geq 0,\ U'(F(t)) - \gamma_0 e^{rt} \leq 0,\ \text{with complementary slackness (c.s.).} \tag{1}$$

where $F(t)$ is the amount of oil extracted and consumed at time t and γ_0 is an undetermined multiplier. Thus, quantity flows according to: $F(t) = U'^{-1}(\gamma_0 e^{rt})$, where U'^{-1} is the inverse of the first derivative of $U(\cdot)$. In addition, the resource stock must get used up either in finite time or asymptotically: $\int_0^\infty F(t)dt = S$, which uniquely determines γ_0 and therefore the time path of extraction and marginal utilities. If we assume that marginal utility is unbounded at zero ($U'(0) = \infty$), then marginal utility must rise forever and the resource stock will only be exhausted in the limit. If we instead assume that marginal utility is bounded at zero ($U'(0) < \infty$), then the resource stock will be exhausted in finite time at the precise instant that the rising marginal utility path reaches its upper bound. This is not only the planner's optimal extraction path but is is the aggregate extraction path that emerges in the competitive equilibrium of a decentralized market.

3.1 Optimal drilling with a finite number of large wells

Suppose instead that oil is distributed among N identical oil wells and that the cost of drilling a well is zero. Like Hotelling's planner, our planner gets utility at each instant from

the flow of oil, a resource in limited supply. However, unlike Hotelling's planner, ours faces a geological constraint: at the moment a well is drilled, its instantaneous flow is X/N, which then decays exponentially to zero at rate λ, so that its instantaneous flow y years after being drilled is $e^{-\lambda y}X/N$. Thus, the total amount of oil in any particular well is $(X/N)/\lambda$, while the aggregate stock of oil across all N wells is $N(X/N)/\lambda = X/\lambda$. Assume that even before the first well is drilled, the initial flow is $F_0 \geq 0$ and that this flow also decays exponentially at rate λ. For purposes of comparison, assume the total volume of oil is the same as in the Hotelling case: $S = (F_0 + X)/\lambda$. The problem of our planner is to decide how many of the N wells to drill at each point over an unbounded time horizon to maximize the present discounted value of utility.

We can solve this complex problem recursively as a sequence of one-variable calculus problems. In each such problem, the planner decides when to drill the next well if R wells remain unopened ($R = 1, \ldots, N$) and the inherited flow is F. We denote the optimal delay or "stall time" until the next well is drilled as $s_R(F)$ and the discounted utility from following the optimal policy when R wells remain to be drilled as $V_R(F)$. After the last well is drilled:

$$V_0(F) = \int_0^\infty e^{-rt} U\left(e^{-\lambda t} F\right) dt. \tag{2}$$

This value is dictated entirely by the inherited flow of oil, since there are zero wells remaining to be drilled.

For any larger number of wells remaining ($R = 1, \ldots, N$), the optimal policy ($s_R(F)$) and corresponding value function ($V_R(F)$) are then defined recursively by the Bellman equation:

$$V_R(F) = \max_{s_R \geq 0} \left\{ \int_0^{s_R} e^{-rt} U\left(Fe^{-\lambda t}\right) dt + e^{-rs_R} V_{R-1}\left(Fe^{-\lambda s_R} + \frac{X}{N}\right) \right\}, \; R = 1, \ldots, N. \tag{3}$$

The first term on the right-hand side is the discounted utility that will accumulate before the next well is drilled because of the inherited flow (F) while the second term is the discounted utility that will accrue after that well is drilled. The latter discounted utility is summarized

by the continuation value function, which depends on the oil flowing immediately after the $N - R^{th}$ well is drilled.

A marginal postponement in drilling the next well has one marginal benefit and two marginal costs. A delay in drilling the next well allows the planner to enjoy marginally longer the utility flow from previously drilled wells. On the other hand, even if the marginal delay did not affect the utility flows in the remainder of the program, postponing these flows is costly at the margin because of discounting; in addition, by delaying, the planner insures that when the next well is drilled, the oil flow from previously drilled wells bequeathed to the future will be smaller. If the planner finds some delay ($s^{*R} > 0$) in opening the next well to be optimal, the marginal benefit from a longer delay must exactly offset the sum of these two marginal costs. If delaying at all has a smaller marginal benefit than these marginal costs, the planner will find it optimal to drill the next well immediately ($s^{*R} = 0$) .

Equation (3) permits calculation of the entire sequence of decision rules: $s_R(F)$ for $R = 1, \ldots, N$. Once this sequence is obtained, one can work forward from the initial flow F_0 to determine when each well should be drilled.

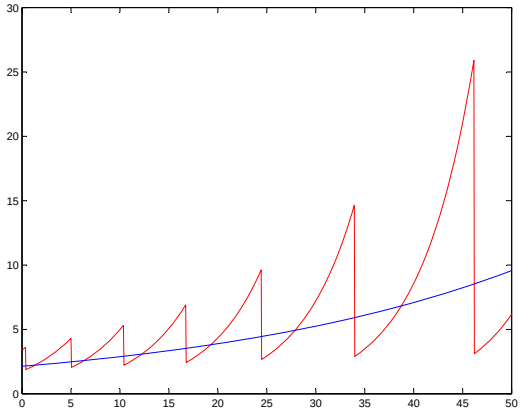
Since $s_R(F) = 0$ is possible, our formulation permits the drilling of two or more wells at the same time. In the most extreme case, it is optimal to drill *every* one of the N wells at the outset. To illustrate, suppose $U(F) = \alpha F^{\beta}$ for $\alpha > 0$ and $\beta \in (0, 1)$. In this case, it is optimal to drill all N wells immediately whenever $\frac{1}{U'(F(t))} \frac{dU'(F(t))}{dt} = \lambda(1 - \beta) < r$. As one works backward, every decision rule is the same: $s_R(F) = 0$ for $R = 1, \ldots, N$.

Typically, however, it will be optimal to delay in drilling at least some wells. In all such cases, the flow jumps up by X/N (or an integer multiple) and the marginal utility jumps down every time a new well is drilled. Consequently, the planner's path of marginal utility inevitably exhibits a saw-tooth pattern. Figure 5 plots the time paths of marginal utility for a particular functional form and parameterization with $N = 20$ and $N = 40$ (in red) along with the standard Hotelling price path (in blue). [16]

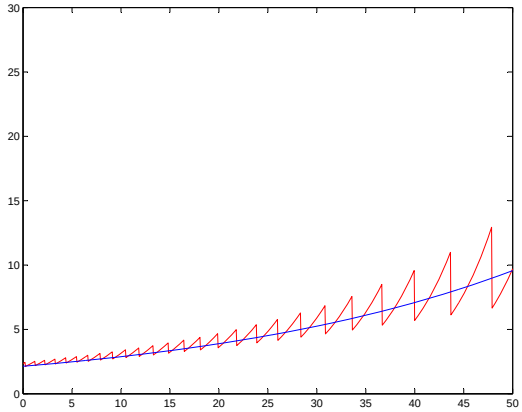
[16] We have shown for this same functional form that the planner's problem cannot be sustained as a price-taking equilibrium. No other candidate equilibrium can exist or it would violate the first welfare theorem.

Figure 5: Sawtooth marginal utility path of our planner compared to smoothly rising path of Hotelling's planner

(a) $N = 10$ wells



(b) $N = 40$ wells



Note: Red lines show marginal utility path of a planner who cannot control the flow from wells once they are tapped. Blue lines show marginal utility path of Hotelling's planner. Total resource stock is the same in all cases. Both figures assume a discount rate of $r = 0.03$, constant elasticity utility parameters of $\beta = 0.1$ and $\alpha = 1$, total resource stock of $S = 1$, initial oil flow of $F_0 = 0$, and rate of decay in oil flow of $\lambda = 0.2$. Initial flow from each well is given by $X = S\lambda/N$.

Figure 5 was constructed assuming a particular utility function with an unbounded $U'(0)$ and zero drilling cost. However, a saw-tooth pattern would typically arise even if we altered these two assumptions. In reality, the oil price does not drop noticeably every time a new well is drilled because even at its peak the flow from any one well is negligible when compared to the aggregate flow of oil from the millions of producing wells around the world. Thus, no matter how we characterize drilling costs and $U'(0)$, we will want to focus on the limiting case where the finite reserves of oil are distributed over an infinite number of infinitesimally small wells.

Intuitively, drilling times in the planner's problem correspond to discrete jumps down in price and therefore local minima in the revenue that a private owner can earn from drilling.

3.2 Optimal drilling with a continuum of small wells

In this case, the saw-tooth pattern disappears and the flow of oil the planner (or the market) generates over time is continuous. After setting up the problem, we derive conditions that necessarily hold at any optimum. This leads to our fundamental result for how drilling revenues and costs should evolve at the optimum or in the associated competitive equilibrium.

3.2.1 Planner's problem and definitions

The planner's problem is given by:

$$\max_{a(t)} \int_{t=0}^{\infty} e^{-rt} \left[U(F(t)) - D(a(t)) \right] dt \tag{4}$$

subject to

$$\dot{F}(t) = a(t)X - \lambda F(t) \tag{5}$$

$$F(0) = F_0 + A(0)X \tag{6}$$

$$\dot{R}(t) = -a(t) \tag{7}$$

$$R(0) = R_0 - A(0), \tag{8}$$

where $a(t)$ is the rate at which new wells are drilled at time t, $F(t)$ is the amount of oil flowing, and $R(t)$ is the amount of wells that remain untapped. The instantaneous utility derived from oil flow is given by $U(F(t))$, where $U(\cdot)$ is increasing and weakly concave. We assume $U(0) = 0$. The total instantaneous cost of drilling wells at rate $a(t)$ is given by $D(a(t))$, where $D(\cdot)$ is increasing and convex. We denote the derivative of the total cost function as $d(a(t))$ and assume that $d(0) \geq 0$. Both flow utility and drilling costs are discounted continuously at rate r.

Condition (5) is the equation of motion governing oil flow $F(t)$—a state variable, since in our problem, unlike Hotelling's, the planner cannot directly control the flow of oil from

existing wells. Instead, the flow from existing wells decays exponentially toward zero at rate λ (the second term), while the planner can only indirectly affect aggregate flow through her rate of new drilling (the first term). This treatment of $F(t)$ as a state variable implicitly assumes that the maximum flow constraint on production is always binding. We focus on this case because, as shown in section 2, it is empirically relevant: expected prices are never observed to rise at a rate greater than r for a sufficiently long period of time so as to make deferring production optimal.[17]

Condition (7) is the equation of motion governing the stock of untapped wells $R(t)$, which declines one-for-one with the rate of drilling. Conditions (6) and (8) are the constraints on initial oil flow and remaining wells. The planning period begins with a continuum of untapped wells of measure R_0 and exogenous oil flow inherited from previously tapped wells of F_0. If the drilling cost is strictly convex, the planner would never find setting up a mass of wells at $t = 0$ to be optimal. When the drilling cost is linear, however, such "pulsing" behavior may be optimal. To take account of it, we distinguish between the exogenous stock of wells remaining to be drilled at the outset of the planning period (R_0) and the endogenous stock of wells remaining if the planner drilled a mass ($A(0)$) of wells at the outset ($R(0) = R_0 - A(0)$).[18] For the same reason, we distinguish between the exogenous flow inherited at the outset of the planning period (F_0) and the endogenous flow that the planner would create if she drilled a mass ($A(0)$) of wells at the same time ($F(0) = F_0 + A(0)X$). The total amount of oil in untapped wells is given by $R(t)X/\lambda$, so that the total amount of oil underground at the outset of the planning period is given by $S = (F_0 + R_0X)/\lambda$.

[17]It is possible to devise examples in which the planner would want to produce below the constraint on the equilibrium path. However, these examples require substantial convexity in the demand curve for oil in order to yield a sufficiently rapid and sustained expected rise in price. Moreover, the planner (or firms) will not want to defer production at any time at which drilling is optimal. Future versions of this paper will discuss incentives to produce below the capacity constraint at greater length.

[18]Later, we use $A(t)$ to denote cumulative drilling through time t. For the case of a finite number of wells considered above, $A(t)$ is constrained to be a step function in time with N steps of height R_0/N—or steps that are integer multiples of R_0/N and correspondingly fewer steps.

3.2.2 Necessary conditions

Following Léonard and Long (1992), the current-value Hamiltonian of this maximization problem is given by:

$$H = U(F(t)) - D(a(t)) + \theta(t)[a(t)X - \lambda F(t)] + \gamma(t)[-a(t)], \tag{9}$$

where $\theta(t)$ and $\gamma(t)$ are the co-state variables on the two state variables $F(t)$ and $R(t)$.

Necessary conditions are given by:

$$a(t) \geq 0,\ \theta(t)X - d(a(t)) - \gamma(t) \leq 0, c.s. \tag{10}$$

$$\dot{F}(t) = a(t)X - \lambda F(t) \tag{11}$$

$$F(0) = F_0 + A(0)X \tag{12}$$

$$\dot{\theta}(t) = -U'(F(t)) + (r + \lambda)\theta(t) \tag{13}$$

$$\dot{R}(t) = -a(t) \tag{14}$$

$$R(0) = R_0 - A(0) \tag{15}$$

$$\dot{\gamma}(t) = r\gamma(t) \tag{16}$$

We begin by interpreting condition (10). On the left-hand side, $\theta(t)X$ is the discounted revenue or utility that an additional well drilled at time t will generate over its lifetime, discounted back to time t, while $d(a(t))$ is the marginal cost (or rental rate) of drilling a well at time t when $a(t)$ other wells are being drilled. Thus, the left-hand side is the marginal profitability of drilling a well at time t. On the right-hand side, we simply have $\gamma(t)$. This co-state variable can be interpreted as the shadow value (in current value terms) of having marginally more than $R(t)$ wells remaining at time t. Note that condition (16) implies that $\gamma(t) = \gamma_0 e^{rt}$, where $\gamma_0 \geq 0$ is a constant. Thus, when $a(t) > 0$, conditions (10) and (16) together imply that:

$$\theta(t)X - d(a(t)) = \gamma_0 e^{rt}. \tag{17}$$

Intuitively, whenever drilling occurs ($a(t) > 0$), the net marginal value of drilling *wells* must rise at the rate of interest so that every well, no matter when it is drilled, yields the same net payoff in present-value terms. This makes perfect sense, for if a planner (or well owner) could earn strictly higher profits by drilling a given well at a different time, then she would surely do so. Condition (10) is analogous to the standard Hotelling rule, which states that the net marginal value of extracting *barrels* rises at the rate of interest so that every barrel extracted yields the same net payoff in present-value terms. Since this is a drilling problem, however, the Hotelling-like intuition applies to wells, not barrels.

Equation (17) relates the discounted stream of drilling revenues to marginal drilling costs when $a(t) > 0$. To see the implication of this condition for the time path of oil prices, assume $a(t) > 0$, differentiate condition (10) with respect to time, and solve for $\dot{\theta}(t)$ to obtain:

$$\dot{\theta}(t) = \frac{r\gamma(t) + d'(a(t))\dot{a}(t)}{X}. \tag{18}$$

Now substitute for $\theta(t)$ and $\dot{\theta}(t)$ in condition (13) using the marginal condition in (10) and condition (18). Assuming that $a(t) > 0$, this yields our other fundamental equation:

$$U'(F(t)) - \frac{(r+\lambda)d(a(t))}{X} + \frac{d'(a(t))\dot{a}(t)}{X} = \frac{\lambda\gamma_0}{X}e^{rt}. \tag{19}$$

On the right-hand side, we have the shadow value of wells ($\gamma_0 e^{rt}$) divided by the total amount of oil stored in a unit mass of untapped wells (X/λ). On the left-hand side, the first term is the marginal utility (or price) of oil. The second term is the amortized, per-barrel marginal cost of drilling a well at time t.[19] The third term captures the opportunity cost of drilling now versus waiting. When $\dot{a}(t) < 0$, drilling activity and marginal drilling costs are

[19]To clarify, consider the special case of constant marginal drilling costs: $d(a(t)) = \bar{d}$. If the planner drills one well (or rather, she marginally increases the rate of drilling) she gets a marginal increase in oil flow h instants later of $Xe^{-\lambda h}$. If each barrel of flow has imputed cost of c at that time, then at the time the well is drilled, the flow at h would have imputed cost $cXe^{-(r+\lambda)h}$. Since such flows continue indefinitely, we want to find c such that $\bar{d} = \int_{h=0}^{\infty} cXe^{-(r+\lambda)h}dh$. Integrating, we get $\bar{d} = cX/(r+\lambda)$. Solving for c we conclude: $c = \bar{d}(r+\lambda)/X$.

falling over time. Thus, drilling immediately incurs an additional opportunity cost relative to delaying.

Condition (19) holds whenever drilling occurs. But when would the planner choose not to drill and allow oil flow to decay exogenously? There are several such scenarios. For example, if inherited oil flow at the outset of the planning period were very high, the planner might postpone in drilling the first well. Or, if during a conjectured interval of incessant drilling, condition (19) implied a decline in oil flow more rapid than was geologically feasible for a non-negative rate of drilling, then drilling would pause. Or finally, if marginal utility were bounded, then the revenue from drilling would also be bounded, implying that the last well would need to be drilled in finite time. None of these scenarios would arise in the canonical Hotelling model, whose planner controls the flow of oil directly.

4 Can the Hotelling rule be saved?

Yes, but only under implausible assumptions. There are two reasons why a planner with our drilling technology would be unlikely to generate Hotelling's extraction path. First, in our model, previously drilled wells will flow forever. Thus, price cannot rise at the rate of interest whenever oil is flowing, as Hotelling's path requires, unless marginal utility is also able to rise forever. Second, in our model, the incentive to drill in a given period depends, in part, on the cost of drilling in *other* periods, as captured by the $d'(a(t))\dot{a}(t)$ term in condition (19), which arises due to the convexity of drilling costs. One implication is that oil flow in our model can *increase* over intervals during which the marginal cost of drilling is falling, whereas the cost of extraction typically increases with production in standard models.

To overcome the first of these threats, we must assume that marginal utility is unbounded: $U'(0) = \infty$. To overcome the second, we must assume that the marginal cost of drilling is constant: $d(a) = \bar{d}$ for all $a \geq 0$. In this case, the imputed per-barrel cost of drilling is well-defined and is given by $c = \bar{d}(r+\lambda)/X$ (see previous footnote). We will assume that c is also

the per-barrel extraction cost faced by the Hotelling planner. Given these two assumptions, condition (19) implies that the marginal utility of oil flow minus the per-barrel marginal cost of extraction rises at the rate of interest:

$$U'(F(t)) - c = \frac{\lambda\gamma_0}{X}e^{rt}. \tag{20}$$

So our planner, although she cannot control the oil flow from any existing well, would drill new wells in such a way that the marginal utility of consumption, net of per-barrel extraction costs, rises over time at the rate of interest. Since cumulative usage of oil is the same along this path and Hotelling's path, the two paths must coincide! This is the limiting case suggested by figure 5, for which $c = 0$ and the total amount of oil was divided among an increasing number of wells ($N = 20, 40$).

Note, however, that we have implicitly assumed that the planner starts drilling wells at the outset of the planning period and never stops, so that condition (19) always applies. For this to be the case, however, we need two other conditions to hold. First, the initial level of oil flow cannot be too high, for otherwise the planner would delay in drilling the first well. Second, the flow from drilled wells must decay sufficiently fast, so that the planner is able to achieve the Hotelling path via judicious control of the drilling rate. If either condition fails, then the geological realities constrain the planner to an inferior path.

To illustrate these two conditions, we assume that drilling costs are zero ($c = 0$) and that utility from oil flow takes the constant elasticity form: $U(F) = \alpha F^\beta$ for $\alpha > 0, \beta \in (0, 1)$. We also assume provisionally that $\lambda(1 - \beta) > r$ and that $F_0 \leq rS/(1 - \beta)$, where $S = (F_0 + R_0X)/\lambda$ is the total resource stock. In this case, both conditions hold, and the

optimal program is given by:

$$\theta(t) = \frac{\gamma_0 e^{rt}}{X}, \text{ where } \gamma_0 = \frac{\alpha\beta X}{\lambda}\left(\frac{rS}{1-\beta}\right)^{\beta-1} \tag{21}$$

$$F(t) = \frac{rS}{1-\beta}e^{-\frac{r}{1-\beta}t} \tag{22}$$

$$a(t) = \frac{\left(\lambda - \frac{r}{1-\beta}\right)}{X}F(t) \tag{23}$$

$$\gamma(t) = \gamma_0 e^{rt}. \tag{24}$$

Note that the planning period begins with pulse of drilling such that oil flow totaling $A(0)X = rS/(1-\beta) - F_0$ is immediately added to the inherited flow. Equation (23) implies that $a(t) > 0$ for all $t \geq 0$, since $F(t) > 0$ and since we have provisionally assumed that $\lambda(1-\beta) - r > 0$. It is straightforward to verify that this program satisfies each of the necessary conditions (10)–(16) and is therefore optimal for our planner.[20] Since $a(t) > 0$, it achieves the same discounted utility as Hotelling's planner.

However, our planner cannot always accomplish this feat. Suppose that the rate of decay from drilled wells is too low, with $\lambda(1-\beta) < r$. To achieve the result in (19) that $U'(F(t)) = \frac{\lambda\gamma_0}{X}e^{rt}$, the planner must set $\dot{F}(t) = \frac{rU'(F(t))}{U''(F(t))} = -\frac{rF(t)}{1-\beta}$. But then equation (11) implies that $a(t)X = \lambda F(t) - \frac{rF(t)}{1-\beta} + \lambda F(t)$. Substituting and simplifying we conclude that $a(t)X = \frac{F(t)(\lambda(1-\beta)-r)}{1-\beta}$, which violates nonnegativity of $a(t)$. Intuitively, the planner is geologically constrained to a price path that rises more slowly than the rate of interest. Even if all of the wells are drilled immediately, oil cannot be extracted quickly enough to satisfy Hotelling's rule.

Suppose instead that the inherited oil flow is too high, with $F_0 > rS/(1-\beta)$. In this case, in which it must be that $\lambda(1-\beta) > r$, the planner will choose not to drill any wells initially, and production will decline at rate λ until drilling commences (when $F_t = rS/(1-\beta)$).

[20] Equations (21) and (24) imply that the marginal condition (10) holds with equality. Equations (21) and (22) imply that condition (13) holds. Equation (24) ensures that (16) holds. Finally, equations (22) and (23) ensure that (11) holds.

During this time, price will rise at a rate greater than r, and our planner will therefore not achieve the same utility as Hotelling's planner.[21]

To summarize, when the extraction technology involves drilling wells rather than producing barrels, we should not expect the Hotelling path to be optimal, unless four conditions hold: (1) the marginal cost of drilling a well is constant, (2) marginal utility is unbounded at zero, (3) the inherited rate of oil flow is not too high, and (4) the decay in flow from drilled wells is sufficiently fast. While the latter two conditions seem reasonable (given that new wells are constantly being drilled in the real world) the former two conditions are not tenable. Our analysis of the Texas data shows clearly that marginal costs rise with the rate of drilling, while the viability of alternative fuels at *current* oil prices argues against an unbounded oil price.

5 Optimal drilling with increasing marginal costs

In this section, therefore, we explore the implications of positive and increasing marginal drilling costs, assuming throughout that the marginal utility of oil is bounded. In a market context, marginal utility equals price, and the marginal cost of drilling equals the rental rate on drilling equipment; we therefore interpret most of our results as market outcomes. We begin by considering the case of an individual oilfield that is small relative to the global market, so that the path of oil prices is exogenous but the local market for renting drilling rigs clears at each instant. We then endogenize the oil price path and focus our attention on two cases that allow for analytical tractability. First, we consider a setting in which scarcity rents are negligible, facilitating an analysis of how unanticipated demand shocks affect oil prices, drilling activity, and production. We then study a setting that includes scarcity rents

[21] If the planner can produce below the production constraint, the standard Hotelling path can still be achieved. However, this possibility requires that drilling is costless. Otherwise, it can be shown that price would initially rise at the rate of interest (while production is below the constraint and drilling is zero), would then rise faster than the rate of interest (after the constraint starts to bind while drilling remains at zero), and finally would rise more slowly than at the rate of interest (after drilling turns positive with the constraint continuing to bind).

and assumes that marginal drilling costs are constant up to a capacity constraint.

5.1 Exogenous oil prices

As a preliminary step to analyzing the full competitive equilibrium, we examine industry supply in isolation. This case is particularly relevant for interpreting drilling behavior in a small, local region, such as Texas, and for describing how that behavior responds to both anticipated and unanticipated changes in the path of world oil prices. We assume that identical extractors face given time paths for oil prices $P(t)$ and drilling rig rental rates $\rho(t)$. We close the model on the input side by assuming a competitive local market for drilling rigs and that this market clears. These assumptions are analogous to the "small country" assumptions in international trade models of an exogenous world output price and endogenous local input prices that clear domestic factor markets.

Intuitively, well owners must decide when to drill based on their forecast of future oil prices ($P(t)$) and drilling rig rental rates ($\rho(t)$). If wells earn strictly positive discounted net profits, then every well will be drilled at some point; no well will remain untapped indefinitely. Since well owners all face the same oil price path and rig rental rate path, they agree in their rankings of the optimal time to rent drilling equipment to open their wells. Rental rates must therefore adjust so that the owners of the wells are *indifferent* about when they drill. Otherwise, all well owners would attempt to rent drilling rigs at the same time and the rental market would not clear. The same intuition applies when we consider the case of endogenous oil prices below.

Industry supply of oil will maximize the discounted profits of the well owners. Thus, we can deduce the path of drilling ($a(t)$) that is most profitable by defining $U(F(t)) = P(t)F(t)$ and $D(a(t)) = \rho(t)a(t)$ and by reconsidering the necessary conditions in (10)-(16).[22] The $\theta(t)X$ term in necessary condition (10) equals the value of the gross revenue stream that

[22]Note in this formulation that utility depends directly on time: $U(F(t), t)$. This could also be the case more generally. For example, incomes may be rising or population may be growing exogenously over time. To avoid complicating the notation, we suppress any implicit time dependence.

results from drilling a well at t, discounted back to that date:

$$\theta(t)X = \int_{y=t}^{\infty} P(y)Xe^{-(r+\lambda)(y-t)}dy, \tag{25}$$

To verify this result, note that differentiating equation (25) with respect to time yields necessary condition (13) for the case where $P(t) = U'(F(t))$. Meanwhile, the $d(a(t))$ term in (10) is the rental price for drilling rigs at time t, where we interpret $\rho(t) = d(a(t))$ as the upward-sloping rig rental supply curve. In equilibrium, well owners are indifferent about when they drill their wells and all earn the same discounted profits, with aggregate industry profits given by the shadow value of undrilled wells: $\gamma_0 R_0$. As we will see, exogenous changes that reduce γ_0 without altering R_0 reduce extractor wealth proportionately. Since the price of oil is exogenous, drilling incentives are unaffected by the oil flow inherited from previously drilled wells; all that matters is the measure of wells remaining to be drilled (R_0). Thus, in what follows, we ignore the inherited flow and focus on drilling and production from the remaining wells.

Suppose that well owners expect the oil price to remain forever at its current level ($\bar{P}$). Then expression (25) implies that $\theta(t)X = \frac{\bar{P}X}{r+\lambda}$, which we assume is sufficiently high that drilling is profitable $\left(\frac{\bar{P}X}{r+\lambda} > d(0)\right)$. During the interval when wells are being drilled, condition (17) then implies that the following equation must hold:

$$\frac{\bar{P}X}{r+\lambda} - \gamma_0 e^{rt} = d(a(t)). \tag{26}$$

A given γ_0 determines the path of rental prices for drilling equipment and hence the path of rigs that would be offered for rent. As we show, a unique γ_0 will induce the supply of the rigs needed to drill the measure of R_0 wells over time. If $\gamma_0 = 0$, the path of rental prices—the left-hand side of (26)—is horizontal and would result in an infinite cumulative supply of rentals, assuming an unbounded time horizon. For any $\gamma_0 > 0$, the left-hand side of (26) decreases over time, reaching $d(0)$ at some date $\hat{T}$. Such a path of rental prices would

induce a finite offering of rig rentals. If γ_0 is larger, the left-hand side is uniformly lower, reaching $d(0)$ sooner. Hence, raising γ_0 generates a uniformly lower path of rental prices and hence a smaller cumulative offering of rigs to rent. If $\gamma_0 \geq \frac{\bar{P}X}{r+\lambda} - d(0)$, the left-hand side reaches $d(0)$ immediately (so $\hat{T} = 0$) and there would subsequently be no drilling rigs offered for rent. Therefore, a unique $\gamma_0 \in \left(0, \frac{\bar{P}X}{r+\lambda} - d(0)\right)$ will induce a path of rental prices such that owners of wells will be able to rent rigs to drill all of the wells.[23]

Having derived the path of drilling activity ($a(t)$) in response to constant oil prices, we can use it in conjunction with the equation of motion for oil flow in (11) to deduce the aggregate supply of oil that results. Drilling activity is intense at first but declines until it ceases altogether at $\hat{T}$. Differentiating equation (11), we conclude that $\ddot{F}(t) = \dot{a}(t)X - \lambda\dot{F}(t)$. Since $\dot{a}(t) < 0, F(t)$ is strictly concave whenever $\dot{F}(t) \geq 0$ and even when $\dot{F}(t)$ is slightly negative. However, for $t > \hat{T}, \dot{a}(t) = 0$ and $\dot{F}(t) < 0$ which implies $\ddot{F}(t) > 0$. Thus, the function $F(t)$ passes through the origin, increases at a decreasing rate from its initial rate of $a(0)X$, reaches a maximum, decreases at an increasing rate but then inflects and ultimately decays exponentially at rate λ.

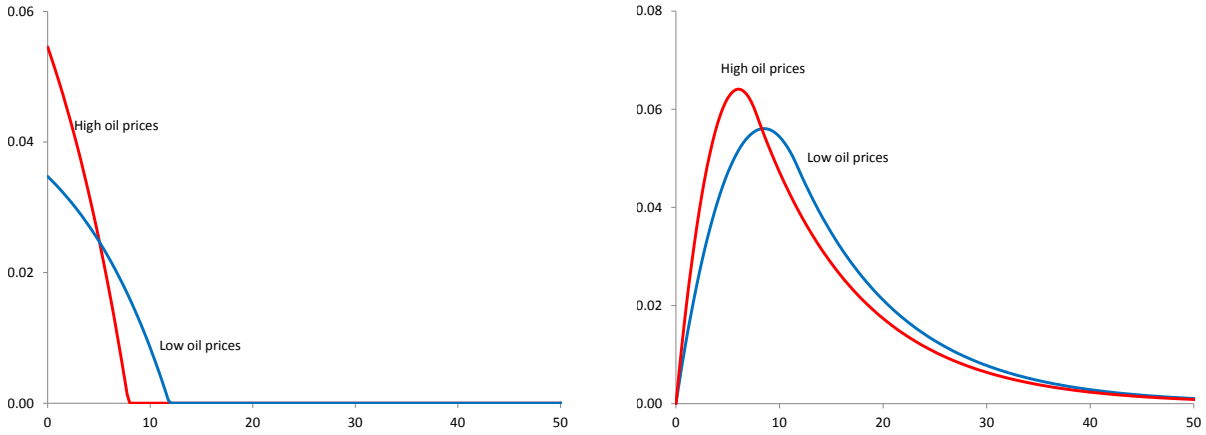
Figure 6 illustrates these results. The hump-shaped "peak oil" profile in panel (b) of this figure is a well-known feature of production on oil fields all across the world. As we have shown, this production profile emerges from a simple model of a competitive oil extraction industry whose only features include a horizontal price path, an upward-sloping local supply function for drilling rigs, and exogenous production from existing wells that decays asymptotically toward zero. We show below that a similar "peak oil" result also emerges in the case of endogenous oil prices.

How do the drilling and production time paths in our local region vary with the exogenous oil price $\bar{P}$? Suppose the oil price expected to persist in the market were higher than that for the case described above. If the γ_0 corresponding to the remaining stock of wells (R_0)

[23]No well owner has an incentive to defer drilling until after $\hat{T}$. Since drilling activity cannot drop below zero, the undiscounted value of drilling after $\hat{T}$ is $\frac{\bar{P}X}{r+\lambda} - d(0)$, making it strictly less profitable in discounted terms than drilling earlier.

Figure 6: Optimal drilling in a local region

(a) Rate of drilling ($a(t)$) **(b)** Oil production ($F(t)$)



Note: This figure illustrates the equilibrium time paths of drilling rates (panel a) and oil production (panel b) for the case of a linear drilling rig supply function; rig rental rates move in tandem with drilling rates. The figure assumes a resource stock of $S = 1$, decay rate in oil flow of $\lambda = 0.1$, initial flow of $X = \lambda S = 0.1$, discount rate of $r = 0.1$, an inverse supply function for drilling rigs with intercept of 0 and slope 10, and low and high oil prices of 1 and 2. See text for details.

did not change at all, then rig rentals would be uniformly higher at every instant and would decline to zero later than before. But then $\int_{t=0}^{\infty} a(t) > R_0$, which cannot occur in equilibrium. Suppose instead that γ_0 increased so much that the initial rental price (the left-hand side of 26) was unchanged at $t = 0$. Then the remainder of the rental price path would be uniformly lower and rigs would be supplied for a shorter period of time. But then $\int_{t=0}^{\infty} a(t) < R_0$, which also cannot occur in equilibrium. Consequently the increase in the oil price must increase *both* γ_0 and the initial drilling rate. The rental price of rigs begins higher but eventually crosses the old equilibrium path and reaches $d(0)$ earlier. As a result, drilling activity starts higher than before but declines more quickly and ceases earlier in response to the higher oil price. Since γ_0 is higher, the increase in the oil price raises the aggregate wealth of the well owners.

Figure 6(a) illustrates these results by comparing drilling rates (and implicitly, rig rental rates) for low and high oil prices. Figure 6(b) compares the resulting flow paths under the

different prices. With a relatively high price, the flow of oil will increase at a steeper rate initially since $a(0)X$ is larger, and by concentrating drilling activity over a shorter time interval, the peak in oil production will occur earlier and be more pronounced. Figure 6(b) illustrates these results.

5.2 Equilibrium dynamics in the absence of scarcity rents

We now close the model by allowing oil prices to be endogenous, and we study the dynamics of oil prices, production, and drilling activity when oil prices are endogenous. To derive analytical results, we assume here unlimited resource stock ($R_0 = \infty$) such that $\gamma(t) = 0$ in condition (10) above. This assumption is an approximation to a world in which scarcity rents are negligible, perhaps because the resource stock is perceived to be vast relative to demand. Consistent with this assumption and its implications, Hamilton (2013) argues that the main driver of oil prices over most of the past century has been the relentless addition of new production capacity to offset the decline in flow from existing fields, leading to stable or falling prices in most periods.

Note that with an unlimited stock of wells, our model reduces to a standard macro model of optimal investment in an industry with convex adjustment costs. Production is a linear function of the stock of productive capital (F), which decays exponentially (at rate λ) but which can be augmented via investment ($a(t)X$). The industry faces a downward-sloping demand curve for its product (with price $P(t) = U'(F(t))$) and a convex investment cost function (with marginal cost $\rho(t) = d(a(t))$). Both revenues and costs are discounted over time (at rate r). In the steady state, all time derivatives are zero ($\dot{a}(t) = \dot{F}(t) = 0$). Solving for steady-state drilling using (11) yields: $a^* = \lambda F^*/X$. Thus, new drilling just offsets the decay in flow from existing wells. Likewise, solving for the steady-state rig rental rate using (19) yields: $U'(F^*)X/(r + \lambda)$, which is equal to the present discounted stream of revenues that a well earns at a constant price of $U'(F^*)$. Thus, well owners earn just enough revenue

to cover the cost of drilling a well. The market is in a long-run equilibrium.[24]

The lack of scarcity in this model—and consequent existence of a steady state—permits analysis of equilibrium dynamics using a two-dimensional phase diagram in $a(t)$ and $F(t)$ space. In particular, we can study the dynamics that follow unanticipated shocks to oil demand. The equations of motion are given by conditions (11) and (19) above, with $\gamma_0 = 0$. Figure 7 presents the $\dot{F}(t) = 0$ and $\dot{a}(t) = 0$ loci (black lines through point B) and corresponding steady state (B). The figure also shows directions of motion within and across the four regions defined by the intersection of these loci (thin gray arrows), as well as the stable arm (saddle path) leading toward the steady state (gray line running southeast through B).[25]

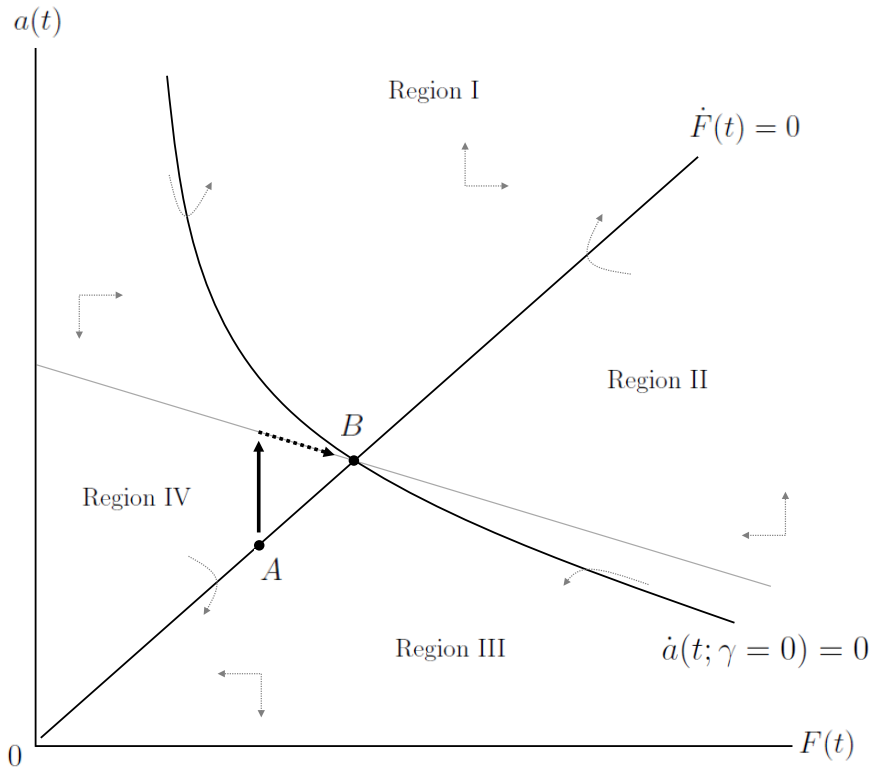
Suppose that oil demand ($U'(F)$) is initially relatively low, so that we are initially in a steady state located at point A. Consider an unanticipated outward shift in oil demand that shifts the $\dot{a} = 0$ locus up and right, yielding a new steady state (B) to the northeast of the old one. The optimal transition dynamics imply jumping up immediately to the new stable arm and then following it down gradually toward the new steady state, as depicted in the figure (thick black arrows). Thus, following a positive demand shock, oil prices, drilling activity, and rig rental rates all spike up on impact, and oil flow begins to increase. Prices then fall gradually over time as production builds up, as do drilling activity and rig rental rates, until we arrive at a new steady state with higher oil prices, flow, drilling, and rig rental rates. For a negative demand shock, the story is reversed, with oil prices, drilling, and rig rental rates all falling sharply on impact and then rising gradually over time to the new steady state. The gradual rise in price can actually proceed at a rate greater than r, at least initially, if the demand decrease is sufficiently large that drilling falls to nearly zero on impact, so that the flow dynamics are dominated by the exponential production decline from drilled wells.[26] Such a steep rise in expected prices actually occurred in winter 1998–1999

[24]Note that we have implicitly assumed that drilling always occurs so that equation (19) always holds.

[25]Since our model with an unlimited resource stock reduces to a standard macro model, we present this phase diagram without derivation, which is available upon request.

[26]Suppose the negative demand shock occurs when the flow rate is given by $\bar{F}$. If the magnitude of the

Figure 7: Phase diagram: infinite stock of oil (special case)



Note: Figure shows the phase diagram corresponding to conditions (11) and (19) with $\gamma(t) = 0$. See text for details.

(see figure 1). Thus, our simple model generates the periods of backwardation and contango following demand shocks—as well as the positive covariance of oil prices, drilling, and rig rental rates—that we observe in the real world.

5.3 Endogenous oil prices when wells are scarce

We finally consider the case of endogenous oil prices when oil is scarce. For tractability, we assume constant marginal drilling costs up to a constraint. We justify this assumption in two ways. First, whereas previous subsections used our theory to replicate key stylized facts

local demand elasticity at $\bar{F}$ is given by $1 - \beta$ and if $\lambda(1 - \beta) > r$, then producing the existing wells at their maximum without any new drilling will cause price to rise more quickly than r.

for the crude oil extraction industry, the purpose of this subsection is to highlight differences between our theory and the standard Hotelling literature, which focuses on deriving hypothetical long-run extraction paths. Our stylized model allows us to derive analytical results that illustrate these differences transparently. In particular, the assumption of constant marginal costs allows us to express costs on a per-barrel basis, facilitating a direct comparison to standard Hotelling results. Second, this special case delivers results that are qualitatively similar to those generated via numerical simulations using more general upward-sloping drilling supply functions.

Given this assumption, the marginal cost of drilling cost takes the following form:

$$d(a(t)) = \bar{d}, \text{ for } a(t) \leq \bar{a}$$
$$d(a(t)) = \infty, \text{ for } a(t) > \bar{a},$$

where $\bar{a}$ is the constraint on drilling. Since the drilling rate is bounded, flow cannot jump instantaneously (or "pulse"); instead, time must elapse for flow to increase. We also assume that the inherited flow of oil is relatively small ($F_0 < \bar{a}X/\lambda$) and that $U''' \geq 0$. With this formulation, necessary condition (10) is replaced by:

$$a(t) \geq 0, \theta(t)X - [\bar{d} + \mu(t)] - \gamma_0 e^{rt} \leq 0, c.s. \tag{27}$$

The shadow price on the capacity constraint, denoted $\mu(t)$, is zero whenever the drilling rate is below the constraint ($a(t) < \bar{a}$); whenever $\mu(t)$ is strictly positive, drilling must be at its maximum feasible rate ($a(t) = \bar{a}$). Thus, the full marginal cost of drilling is given by $\bar{d}+\mu(t)$, which we can interpret as the price to rent a drilling rig in a competitive market, inclusive of the rent on scarce capacity.[27]

[27]Feng, Zhao and Kling (2002) study the optimal time path of carbon sequestration using a setup that is mathematically similar to ours, with the stock of carbon in the atmosphere playing the role of our oil flow from existing wells and the fixed amount of land suitable for carbon sequestration activities playing the role of our fixed stock of wells to be drilled. Like us, they derive analytical results for the case of constant marginal costs (of sequestration) up to a period capacity constraint.

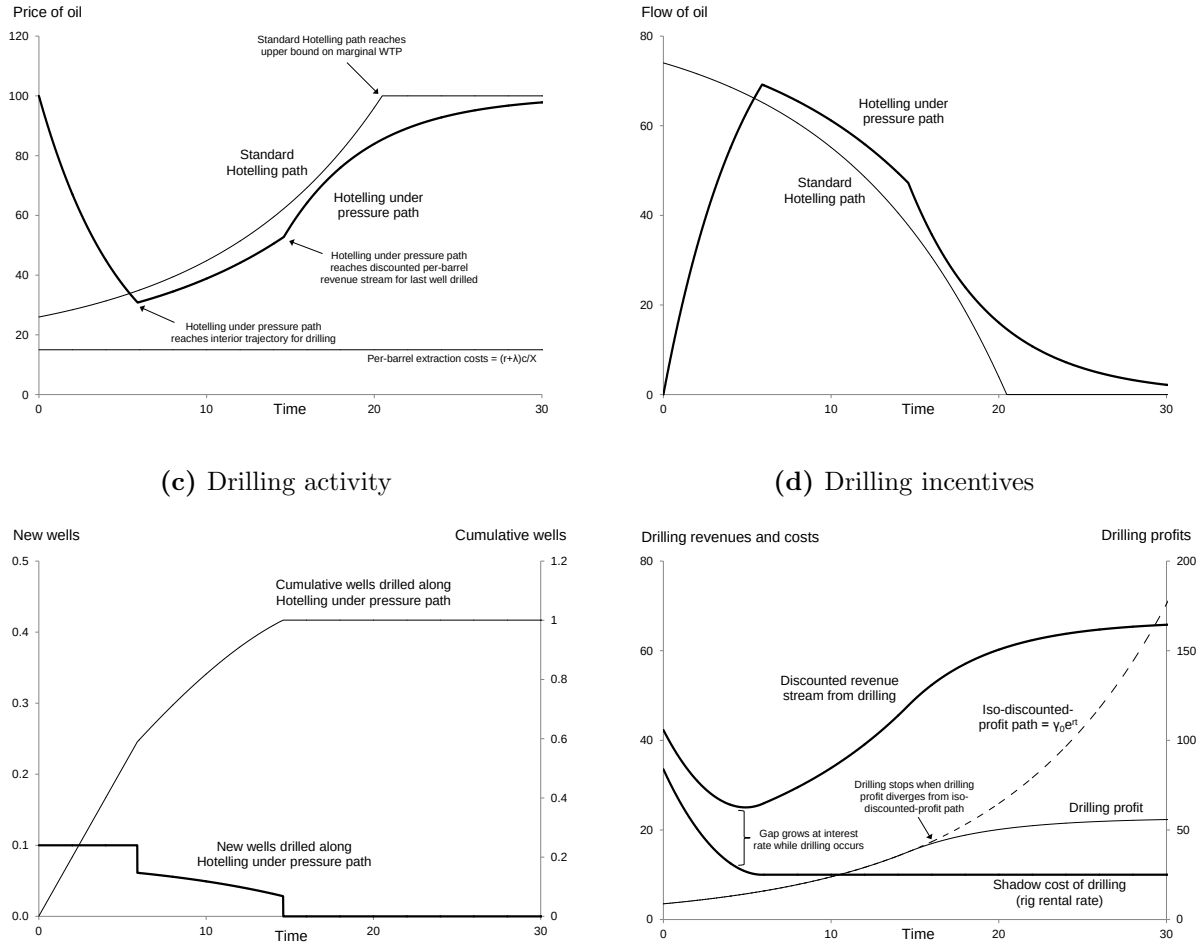
The optimal program consists of three intervals, and the boundaries between them are determined endogenously. During the first interval, drilling is set at the maximum feasible rate ($a(t) = \bar{a}$, $\mu(t) > 0$), with the flow of oil rising at a decreasing rate ($\dot{F}(t) > 0, \ddot{F}(t) < 0$). During the second interval, drilling proceeds at a slower rate ($a(t) \in (0, \bar{a})$, $\mu(t) = 0$). with the flow of oil falling at an increasing rate ($\dot{F}(t) < 0, \ddot{F}(t) < 0$) and drilling decreasing over time ($\dot{a}(t) < 0$). Thus, oil flow is kinked at the first boundary, implying a discrete drop in the rate of drilling at the boundary. Finally, during the third interval, drilling drops to zero ($a(t) = 0$) and oil flow decays toward zero exponentially ($\dot{F}(t) < 0, \ddot{F}(t) > 0$). Thus, oil flow is also kinked at the second boundary. Overall, the extraction path is characterized by a peak in production, implying an initial period of falling prices as production ramps up, followed by a period of rising prices as drilling slows, and finally a period of exogenously rising prices as drilled wells all drain to exhaustion.[28]

In appendix B, we establish these properties formally and describe an algorithm for determining the boundaries between time intervals and the shadow value on the resource constraint (γ_0). If the stock of wells (R_0) is higher, then the corresponding shadow value falls, resulting in a longer first interval, uniformly higher flow during the second interval, and later termination of the second interval.

Figure 8 depicts a simulation of this model for the case of quadratic utility (satisfying $U'(0) < \infty, U''' \geq 0$) and assuming $F_0 = 0$. The reader can verify that this figure illustrates the properties described above for the time paths of oil prices (panel a), production (panel b), and drilling (panel c). For comparison, the figure also depicts Hotelling's time path for prices and production (in panels a and b). Lastly, the figure depicts our paper's key necessary condition for optimality, which is that the discounted revenue stream minus marginal cost of drilling a well must rise at the rate of interest while drilling occurs (panel d). Since revenues in this case are constrained by the upper bound on oil prices, drilling must eventually cease.

[28] In a more general model in which $d'(a(t))$ were continuous and strictly greater than zero, there would not be kinks in oil flow nor discrete changes in drilling rates. However, drilling would still fall monotonically over time, and production would still exhibit a peak.

Figure 8: Hotelling and Hotelling-under-pressure paths



Note: This figure illustrates the optimal time paths of drilling, oil production, and prices for the case of a linear demand curve and constant marginal drilling costs up to a capacity constraint. Figure assumes $\lambda = 0.2$, $X = 200$, $R = 1$, and $S = RX/\lambda = 1000$ for the oil reserve, an intercept of $a = 100$ and slope of $b = -1$ for the linear demand function, a marginal drilling cost of $c = 10,000$ with capacity constraint of $\bar{a} = 0.1$, initial oil flow of $F_0 = 0$, and discount rate of $r = 0.1$. See text for details.

Throughout this section, we have assumed a bounded oil price ($U'(0) < \infty$) and a constrained drilling rate ($\bar{a} < \infty$). If the oil price were unbounded, then the equilibrium would be similar but without the final interval: drilling would never cease entirely. If instead there were no constraint on drilling, then the first phase would disappear. At the first instant, wells would be drilled at an *infinite* rate (a "pulse") so that equation (19) would

hold immediately at $t = 0$ and would continue to hold until drilling ceased.

6 Conclusion

Standard Hotelling models take for granted that extractors have precise control over the rate at which production flows to market. Our analysis of crude oil drilling and production in Texas shows this assumption to be inconsistent with the technology and cost structure of the crude oil extraction industry—and presumably the natural gas industry as well. Oil is not extracted barrel-by-barrel. Instead, extractors drill wells that then flow to exhaustion at a rate that is geologically constrained.

We develop a new model of exhaustible resource extraction that accommodates these important features of the crude oil extraction industry. Oil extractors in our model maximize wealth by choosing when to drill wells, but the flow from these wells decays exogenously. With a finite number of wells, the social planner is constrained to a saw-tooth price path that is inconsistent with price-taking behavior. For the more realistic case of a large number of small wells, however, our model replicates several salient facts observed in the real world: (1) production from pre-existing wells steadily declines over time and does not respond to oil price shocks; (2) drilling of new wells and drilling rig rental rates strongly co-vary with oil prices; (3) local oil-producing regions and fields exhibit production peaks; and (4) expected future oil prices can be backwardated after positive demand shocks and can rise faster than the interest rate (temporarily) following negative demand shocks. Standard Hotelling models cannot explain any of these facts, and the literature that allows for "set up costs," though able to explain production peaks and price backwardation, fails to rationalize why production from existing wells can decline over time but at the same time can be insensitive to price shocks. Such models also do not permit an expectation that prices will rise faster than the rate of interest.

Our model could be extended in several logical ways. We assume a single continuum

of homogenous wells, all located within the same, static market for drilling rigs. Instead, we could divide the continuum of wells across two or more regions—say Texas and Alaska, or onshore and offshore—each with its own supply of drilling rigs. Since different locations typically have their own geological features, it would also be natural to consider variation in drilling costs, production decline rates, and resource stocks across regions—or across individual wells within the same region. Lastly, while the stock of drilling rigs and crews in a given region is fixed in the short run, this stock can change over time, as new rigs are built and crews are trained, as old rigs are scrapped and workers retire, or as existing rigs and crews are moved from one region to another. These drilling industry dynamics could be incorporated to enrich the model. We leave these extensions to future work.

References

Alquist, Ron and Lutz Kilian, "What Do We Learn From the Price of Crude Oil Futures?," *Journal of Applied Econometrics*, June/July 2010, *25* (4), 539–573.

Berck, Peter and Michael Roberts, "Natural Resource Prices: Will They Ever Turn Up?," *Journal of Environmental Economics and Management*, 1996, *31*, 65–78.

Black, Geoffrey and Jeffrey T. LaFrance, "Is Hotelling's Rule Relevant to Domestic Oil Production?," *Journal of Environmental Economics and Management*, 1998, *36*, 149–169.

Dahl, Carol and Mine Yucel, "Testing Alternative Hypotheses of Oil Producer Behavior," *The Energy Journal*, 1991, *0* (4), 117–138.

Darcy, Henry, *Les Fontaines Publiques de la Ville de Dijon* 1856.

Devarajan, Shantayanan and Anthony C. Fisher, "Hotelling's "Economics of Exhaustible Resources": Fifty Years Later," *Journal of Economic Literature*, 1981, *19* (1), 65–73.

Feng, Hongli, Jinhua Zhao, and Catherine L Kling, "The time path and implementation of carbon sequestration," *American Journal of Agricultural Economics*, February 2002, *84* (1), 134–149.

Gaudet, Gerard, "Natural resource economics under the rule of Hotelling," *Canadian Journal of Economics/Revue Canadienne d'Economique*, 2007, *40* (4), 1033–1059.

_ **and Stephen Salant**, "The Hotelling Model with Multiple Demands," in Robert Halvorsen and David Layton, eds., *Handbook on the Economics of Natural Resources*, Edward Elgar, 2014.

Griffin, James M, "OPEC Behavior: A Test of Alternative Hypotheses," *American Economic Review*, December 1985, *75* (5), 954–63.

Halvorsen, Robert and Tim R. Smith, "On Measuring Natural Resource Scarcity," *Journal of Political Economy*, 1984, *92* (5), 954–964.

Hamilton, James D., "Understanding Crude Oil Prices," *Energy Journal*, 2009, *30* (2), 179–206.

_ , "Oil Prices, Exhaustible Resources, and Economic Growth," in Roger Fouquet, ed., *Handbook of Energy and Climate Change*, Cheltenham, United Kingdom: Edward Elgar Publishing, 2013.

Hogan, William W., "A Dynamic Putty–Semi-putty Model of Aggregate Energy Demand," *Energy Economics*, 1989, *11* (1), 53–69.

Holland, Stephen P., "Modeling Peak Oil," *The Energy Journal*, 2008, *29* (2), 61–79.

Hotelling, Harold, "The Economics of Exhaustible Resources," *The Journal of Political Economy*, 1931, *39* (2), 137–175.

Hyne, Norman J., *Nontechnical Guide to Petroleum Geology, Exploration, Drilling, and Production*, 2nd edition ed., Tulsa, Oklanhoma: PennWell Corporation, 2001.

Jones, Clifton T., "OPEC Behaviour Under Falling Prices: Implications for Cartel Stability," *The Energy Journal*, 1990, *0* (3), 117–130.

Kellogg, Ryan, "The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling," *American Economic Review*, Forthcoming.

Kilian, Lutz, "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review*, 2009, *99* (3), 1053–1069.

Kline, Patrick, "Understanding Sectoral Labor Market Dynamics: An Equilibrium Analysis of the Oil and Gas Field Services Industry," November 2008. Unpublished Manuscript.

Léonard, Daniel and Ngo van Long, *Optimal Control Theory and Static Optimization in Economics* number 9780521331586. In 'Cambridge Books.', Cambridge University Press, November 1992.

Livernois, John R. and Russell S. Uhler, "Extraction Costs and the Economics of Nonrenewable Resources," *Journal of Political Economy*, 1987, *95* (1), 195–203.

Mason, Charles F. and Klaas van't Veld, "Hotelling Meets Darcy: A New Model of Oil Extraction," June 2013. Paper presented at the 20th Annual Conference of the European Association of Environmental and Resource Economists (Toulouse, France).

Miller, Merton H. and Charles W. Upton, "A Test of the Hotelling Valuation Principle," *Journal of Political Economy*, 1985, *93* (1), 1–25.

Pindyck, Robert S., "The Optimal Exploration and Production of Nonrenewable Resources," *Journal of Political Economy*, October 1978, *86* (5), 841–61.

Ramcharran, Harri, "Oil Production Responses to Price Changes: An Empirical Application of the Competitive Model to OPEC and Non-OPEC Countries," *Energy economics*, 2002, *24* (2), 97–106.

Rao, Nirupama S., "Taxation and the Extraction of Exhaustible Resources: Evidence from California Oil Production," April 2010. MIT-CEEPR Working Paper 10-006.

Salant, Stephen W., "The Economics of Natural Resource Extraction: A Primer for Development Economists," *World Bank Research Observer*, February 1995, *10* (1), 93–111.

Slade, Margaret E., "Trends in Natural-resource Commodity Prices: An Analysis in the Time Domain," *Journal of Environmental Economics and Management*, 1982, *9*, 122–137.

_ **and Henry Thille**, "Whither Hotelling: Tests of the Theory of Exhaustible Resources," *Annual Review of Resource Economics*, 09 2009, *1* (1), 239–259.

Smith, V. Kerry, "Natural Resource Scarcity: A Statistical Analysis," *Review of Economics and Statistics*, 1979, *61*, 423–427.

Society of Petroleum Evaluation Engineers, "Survey of Economic Parameters Used in Property Evaluations," 1995.

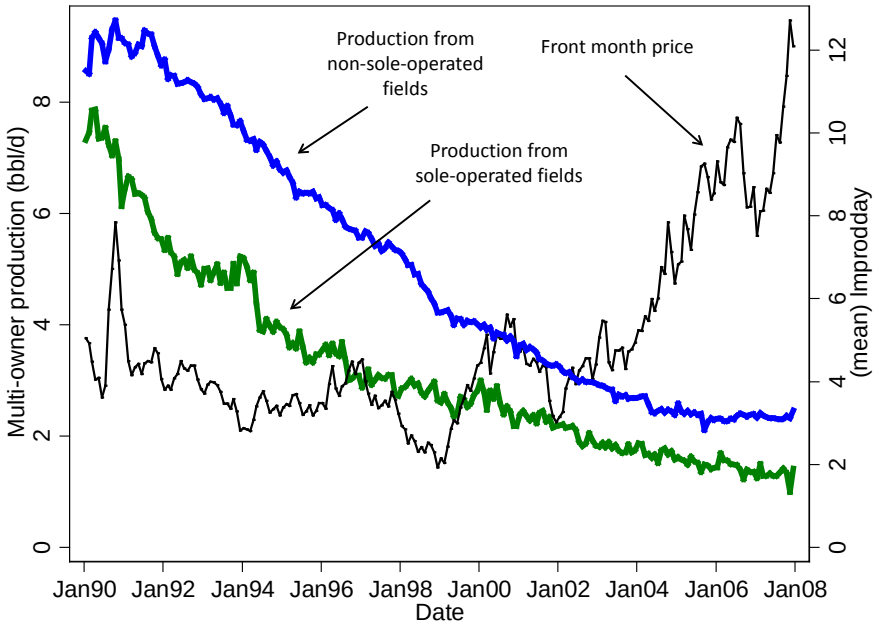
Switzer, Sheldon and Stephen Salant, "Optimal Capacity Expansion by Foresighted Extractors and Producers (Expansion Optimale de Dapacité par des Exploitants Prévoyants d'une Ressource Renouvelable)," in G. Gaudet and P. Lasserre, eds., *Problemes de Ressources Naturelles et Theorie Economique*, University of Laval Press, Laval, Quebec, 1986.

Thompson, Andrew C., "The Hotelling Principle, Backwardation of Futures Prices and the Values of Developed Petroleum Reserves—the Production Constraint Hypothesis," *Resource and Energy Economics*, 2001, *23*, 133–156.

A Ruling out alternative explanations for production's lack of response to price incentives

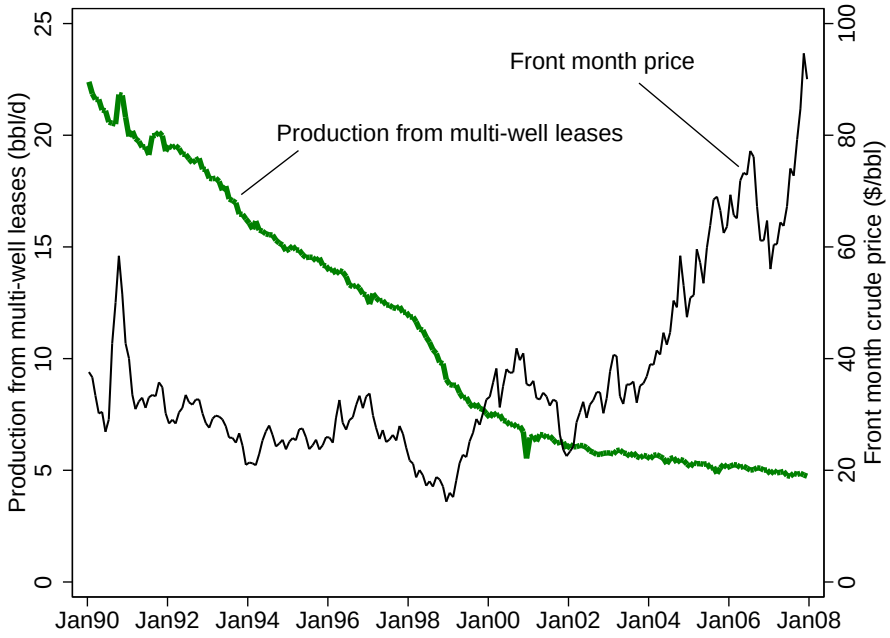
There are several potential alternative explanations for the lack of price response among existing oil wells, which we need to rule out. First, races to oil created by common-pool externalities in oil fields with multiple lease-holders would diminish the incentive that individual lease-holders have to defer production when prices are expected to rise, since much of the deferred production would be extracted by others instead. Figure 9 shows, however, that the long, downward trend in oil production manifests both for oil fields with multiple operators, as well as for oil fields with just a single operator. If anything, fields with multiple operators showed a bigger decline in production around 1998, perhaps because a larger share of leases on these fields have marginal production.

Figure 9: Sole-operated fields versus multiple-operator fields



Note: This figure shows average daily production on fields with multiple operators as well as on fields with just a single operator, along with the spot price of crude oil. See text for details.

Figure 10: Production on multi-well leases

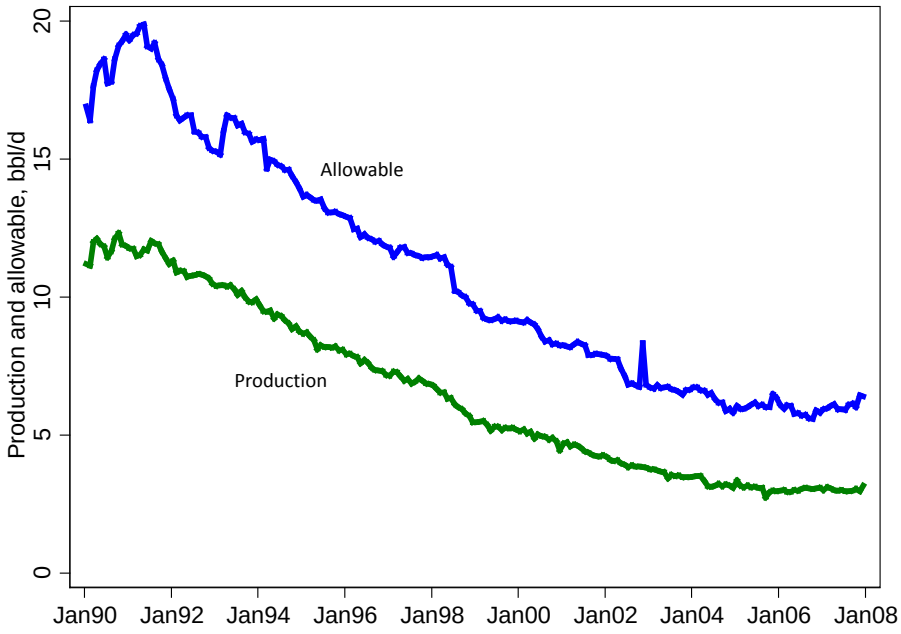


Note: This figure shows average production for leases with multiple wells in our main sample. See text for details.

Second, a condition of many leases is that the lease holder produce oil; a firm that drops lease-level production to zero may therefore risk losing the lease. Figure 10 shows, however, that production from multi-well leases, on which producers may shut in at least some of their wells without risk of losing the lease, does not respond to price signals.

Third, oil production in Texas is subject to maximum allowable production quotas—or "allowables"—as determined by the Texas Railroad Commission. This system dates to the East Texas Oil Boom of the 1930s when races to oil led to overproduction and collapsing world oil prices. Whether originally intended to end the race to oil, or simply to cartelize the Texas oil industry and boost prices, this system persists to this day, and very lease in our data has a monthly allowable, including on fields with just a single operator. One obvious concern is that these maximum production quotas are binding, leading to the lack of price response. Figure 11 shows, however, that the average production for leases in our main sample is well below the average allowable production. Thus, the allowables are not binding

Figure 11: Actual versus allowable production

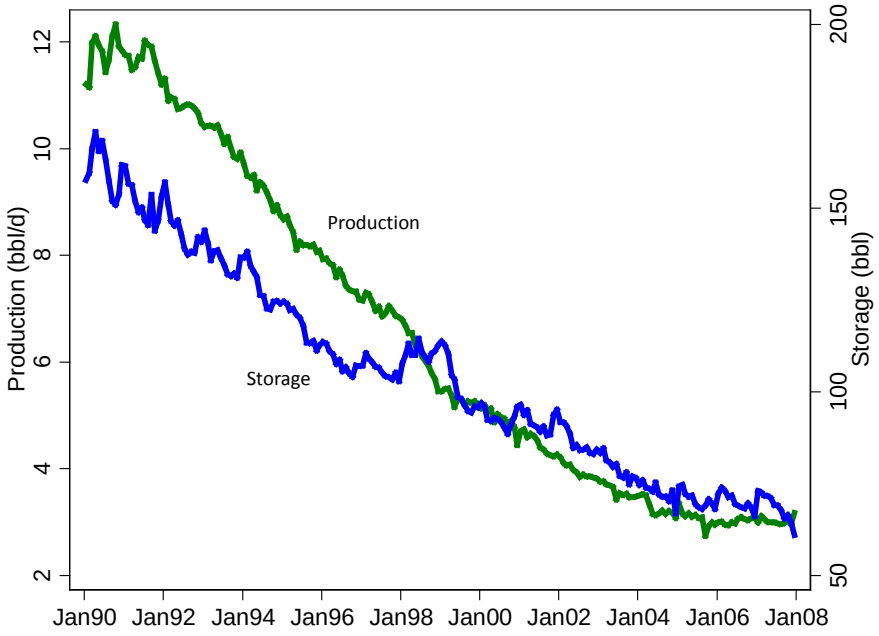


Note: This figure shows average actual production as well as average allowable production for leases in our main sample. See text for details.

and therefore cannot explain the lack of price response that we observe in our data.

Fourth and finally, one possible concern is that the decision makers whose behavior we observe in our data could earn profits by delaying production during periods of extreme contango but do not, perhaps because they are myopic or simply do not understand the potential to lock-in guaranteed profits by deferring production and taking a short position in the futures market. Figure 12 shows, however, that above-ground storage of crude oil on these leases increased notably during the 1998 period of extreme contango. Lease-holders responded to these price incentive by accumulating inventories above-ground, deferring sales—not extraction—to take advantage of the expected increase in prices.

Figure 12: Production and above-ground storage



Note: This figure shows average oil production and average above-ground storage of crude oil on leases in our main sample. See text for details.

B Details for endogenous price model with scarcity

We now examine the case of endogenous oil prices. To make this case analytically tractable, we assume that marginal drilling costs are constant up to a capacity constraint, which is a special case of increasing marginal costs.[29] Marginal drilling costs are given by:

$$d(a(t)) = \bar{d}, \text{ for } a(t) \leq \bar{a}$$

$$d(a(t)) = \infty, \text{ for } a(t) > \bar{a},$$

where $\bar{a}$ is the capacity constraint on the drilling rate. Since the drilling rate is bounded, the rate of oil flow cannot jump instantaneously (or "pulse") as it did in section 4. Time

[29]Feng et al. (2002) study the optimal time path of carbon sequestration using a setup that is mathematically similar to ours, with the stock of carbon in the atmosphere playing the role of our oil flow from existing wells and the fixed amount of land suitable for carbon sequestration activities playing the role of our fixed stock of wells to be drilled. Like us, they derive analytical results for the case of constant marginal costs (of sequestration) up to a period capacity constraint.

must elapse for flow to increase. We assume that the inherited flow of oil is relatively small ($F_0 < \bar{a}X/\lambda$) and that $U''' \geq 0$.

With this formulation, necessary condition (10) is replaced by:

$$a(t) \geq 0, \theta(t)X - [\bar{d} + \mu(t)] - \gamma_0 e^{rt} \leq 0, c.s. \tag{28}$$

The shadow price on the capacity constraint, denoted $\mu(t)$, is zero whenever the drilling rate is below the constraint ($a(t) < \bar{a}$); whenever $\mu(t)$ is strictly positive, drilling must be at its maximum feasible rate ($a(t) = \bar{a}$). Thus, the full marginal cost of drilling is given by $\bar{d}+\mu(t)$. We can interpret $\bar{d}+\mu(t)$ as the price to rent a drilling rig in a competitive market inclusive of the rent on scarce capacity.

The optimal program consists of three intervals. During the first interval, for $t \in [0, \hat{t}]$, drilling is set at the maximum feasible rate ($a(t) = \bar{a}$) with $\mu(t) > 0$ in condition (28). During the second interval, for $t \in (\hat{t}, \hat{T}]$, drilling proceeds at a slower rate ($a(t) \in (0, \bar{a})$) with $\mu(t) = 0$ in condition (28). Finally, during the third interval, for $t \in (\hat{T}, \infty)$, the drilling rate jumps down to zero ($a(t) = 0$) and oil flow decays exogenously. The boundaries between these three time intervals, given by $\hat{t}$ and $\hat{T}$, are determined endogenously.

How do drilling and oil flow evolve during these three intervals? During the first interval, drilling is set at the maximum rate and the flow of oil increases monotonically from its initial level of F_0 at a decreasing rate. To verify that $\dot{F}(t) > 0$ and $\ddot{F}(t) < 0$, note that equation (11) simplifies to the following when $a(t) = \bar{a}$

$$F(t) = \frac{\bar{a}X}{\lambda} + k_0 e^{-\lambda t} \text{ for } t = [0, \hat{t}), \tag{29}$$

where $k_0 = F_0 - \bar{a}X/\lambda < 0$, given our assumption on inherited oil flow. Since $e^{-\lambda t}$ is a strictly decreasing, strictly convex function and $k_0 < 0$, $F(t)$ is a strictly increasing, strictly concave function for $t \in [0, \hat{t})$, as was to be proved.

During the second interval, the flow of oil decreases over time and this decrease grows in

magnitude monotonically as time elapses ($\dot{F}(t) < 0, \ddot{F}(t) < 0.$). In addition, drilling must strictly decrease over time ($\dot{a}(t) < 0$). To establish these conclusions, differentiate condition (19) with respect to time to obtain:

$$U''(F(t))\dot{F}(t) = r\frac{\lambda\gamma_0}{X}e^{rt}, \tag{30}$$

recalling here and throughout that marginal drilling costs are constant. Use (30) to eliminate $\dot{F}(t)$ from the equation of motion in (11). Use (19) to eliminate γ_0, and finally solve for $a(t)X$:

$$a(t)X = \frac{r[U'(F(t)) - \frac{r+\lambda}{X}\bar{d}]}{U''(F(t))} + \lambda F(t). \tag{31}$$

Note that the first term on the right-hand side of equation (31) is negative. Subtracting $\lambda F(t)$ from both sides and recalling equation (11), it is clear that $\dot{F}(t) < 0$. Moreover, our assumption that $U'''(\cdot) \geq 0$ implies that the negative first term on the right-hand side of equation (31) grows in magnitude over time. This implies that $\ddot{F}(t) < 0$ and that $\dot{a} < 0$ for $t \in (\hat{t}, \hat{T}]$, as was to be proved.

During the third and final interval, drilling jumps down to zero. Denote the oil flow at the moment when drilling stops as $\hat{F} = F(\hat{T})$. Oil flow decays exponentially from $\hat{F}$ at the exogenous rate λ. Hence, $\dot{F}(t) < 0$ and $\ddot{F}(t) > 0$ for $t \in (\hat{T}, \infty)$.

How do drilling and oil flow behave at the boundaries? To begin, denote the number of wells drilled up through time t as $A(t)$. Since the drilling rate is bounded below by zero and above by $\bar{a} > 0$, there can be no upward or downward jumps in $A(t)$. The function is continuous although, as we will see, not differentiable everywhere. In the first interval, $A(t)$ rises linearly: $A(t) = \bar{a}t$ for $t \in [0, \hat{t})$. In the second interval,

$$A(t) = \bar{a}\hat{t} + \int_{\hat{t}}^{t} a(s)ds, \tag{32}$$

where $a(t)$ is given by (31) above. In the third interval, $A(t) = A(\hat{T}) = \bar{a}\hat{t} + \int_{\hat{t}}^{\hat{T}} a(s)ds$. The

rate of drilling can jump up or down in this problem because, when $D(a(t)) = \bar{d}a(t)$, the Hamiltonian is linear in the control variable $a(t)$. However, the jumps are finite since $a(t)$ is bounded and cannot produce discontinuities in $A(t)$. They do, however, produce kinks—not only in $A(t)$ but also in $F(t)$. Kinks in both functions must occur at the boundary between the first and second interval ($\hat{t}$) and again at the boundary between the second and third interval ($\hat{T}$).

Now consider what occurs at the first of the two boundaries. Since $\dot{F}(t)$ is strictly positive throughout the first interval and strictly negative throughout the second interval, $F(t)$ must be kinked at the boundary between the two intervals ($\hat{t}$). The left-derivative there is strictly positive while the right-derivative is strictly negative. For this to occur, the rate of drilling must jump down at the end ($\hat{t}$) of the first interval: $\bar{a} > a(\hat{t}^+)$. Since the left-derivative of $A(t)$ at $\hat{t}$ is $\bar{a}$ and the right-derivative is $a(\hat{t}^+)$, $A(t)$ is also kinked at $\hat{t}$.

Consider next what occurs at the second of the two boundaries. Drilling activity jumps down to zero at $\hat{T}$. Since $\dot{A}(t) = a(t)$, there must be a kink at $\hat{T}$ in $A(t)$—in particular, the left derivative of $A(\hat{T})$ is strictly positive and the right derivative is zero. Since $\dot{F}(t) = a(t)X - \lambda F(t)$ and $F(t)$ is continuous, there is a kink at $\hat{T}$ in $F(t)$—its right derivative is negative and its left derivative is larger (strictly positive or at least not as negative).

Figure 8 depicts a simulation of our model under the assumption that $F_0 = 0$ and utility is quadratic (satisfying $U'(0) < \infty, U''' \geq 0$). The reader can verify that this figure illustrates the properties established above.

Finally, how do we determine the optimal program in any given instance? In particular, how do we determine the boundaries between intervals ($\hat{t}$ and $\hat{T}$) and the shadow value on the resource constraint (γ_0)? The first interval ends at $\hat{t}$. Since $\mu(t) = 0$ at the transition, condition (19) holds at the transition, and so the transition time as a function of γ_0 is given implicitly by the following expression:

$$U'\left(\frac{\bar{a}X}{\lambda} + k_0 e^{-\lambda \hat{t}}\right) - \frac{\lambda \gamma_0}{X} e^{r\hat{t}} = \frac{(r+\lambda)\bar{d}}{X}. \tag{33}$$

Meanwhile, it can be shown that oil flow at the moment the last well is drilled, denoted by $\hat{F}$, is independent of γ_0 and is defined implicitly by:

$$\frac{U'(\hat{F}) - \frac{r\bar{d}}{X}}{\lambda} = \int_0^\infty U'\left(\hat{F}e^{-\lambda t}\right) e^{-(r+\lambda)t} dt. \tag{34}$$

Briefly, from the moment the last well is drilled, oil flows exogenously. Thus, the marginal value of oil flow from that moment onward is a deterministic function of oil flow (the right-hand side). At the same time, the necessary conditions combine to yield the marginal value of oil flow prior to that moment as a function of oil flow (the left-hand side). Both conditions must hold simultaneously at the boundary, pinning down oil flow at the moment the last well is drilled.[30]

These observations lead to a straightforward iterative method to determine γ_0, $\hat{t}$, and $\hat{T}$. The procedure is as follows. First, guess some $\gamma_0 \geq 0$. Second, compute the implied flow over the first two intervals. Third, assume that drilling stops when oil flow reaches $\hat{F}$,

[30]Let $\hat{T}$ denote the endogenously chosen time that the last well is drilled. Imposing exogenously that the planner stop drilling at her optimal $\hat{T}$ will obviously not affect her maximized utility. Moreover, we can treat the utility accruing after $\hat{T}$ as a "scrap value" that depends on oil flow $\hat{F}$ at $\hat{T}$. Altering the program to obtain some different oil flow at $\hat{T}$ cannot raise the value of the scrap value at $\hat{T}$ plus the utility accruing before $\hat{T}$. Thus, at $\hat{T}$, the standard endpoint condition for the case of a free state variable and fixed terminal time with a scrap value applies:

$$\theta(\hat{T}) = e^{r\hat{T}} \frac{\partial}{\partial \hat{F}} \int_{s=\hat{T}}^\infty e^{-rs} U(\hat{F}e^{-\lambda(s-\hat{T})}) ds.$$

The right-hand side is the benefit at $\hat{T}$ of endowing the final interval with marginally greater oil flow; the left-hand side is the marginal cost in terms of foregone net utility earned prior to $\hat{T}$. Differentiating, we obtain:

$$\theta(\hat{T}) = \int_{s=\hat{T}}^\infty U'\left(F(\hat{T})e^{-\lambda(s-\hat{T})}\right) e^{-(r+\lambda)(s-\hat{T})} ds.$$

But from equation (19) and equation (28) we can write $\theta(\hat{T})$ as follows:

$$\theta(\hat{T}) = \frac{\bar{d}}{X} + \frac{\gamma_0 e^{r\hat{T}}}{X} = \frac{U'(F(\hat{T}) - \frac{r\bar{d}}{X}}{\lambda}.$$

Thus, eliminating $\theta(\hat{T})$ using these last two equations, we conclude:

$$\frac{U'(\hat{F}) - \frac{r\bar{d}}{X}}{\lambda} = \int_0^\infty U'\left(\hat{F}e^{-\lambda t}\right) e^{-(r+\lambda)t} dt.$$

which is independent of the initial guess of γ_0. Fourth, compute the flow throughout the final interval. Fifth, check whether the cumulative flow over the three intervals matches the total resource stock ($S = (F_0 + R_0 X)/\lambda$). Sixth, stop if the two sums match; otherwise, revise γ_0, recognizing that a higher γ_0 will result in a smaller cumulative oil flow. There can be at most one γ_0 that results from this solution algorithm, and every necessary condition is satisfied by this solution.

Cumulative oil flow is calculated as follows. Oil flow during the first interval is given by equation (29). The first interval ends at $\hat{t}$, which is defined implicitly by equation (33). Since the left-hand side of this equation is strictly decreasing in $\hat{t}$ there can be at most one solution to this equation. For future reference, note that if γ_0 were higher, the first interval would end sooner (smaller $\hat{t}$). Oil flow during the second interval is given by equation (19). Since the left-hand side of this equation is decreasing in $F(t)$, the equation uniquely defines $F(t)$ at every moment during the second interval. For future reference, note that if γ_0 were higher, the flow at every instant during the second interval would be uniformly smaller. Assume that the second interval ends when the flow drops to the $\hat{F}$ threshold defined above. Finally, oil flow during the final interval decays exponentially at rate λ from an initial level of $\hat{F}$, which is independent of γ_0.

Either the cumulative flow of oil over the three intervals matches $S = (F_0 + R_0 X)/\lambda$ or it does not. If the cumulative flow of oil is larger than S, then we have not identified the optimum and should raise γ_0. The first interval (during which drilling is set to the constraint) will be shorter, while the second interval will have uniformly lower oil flow and will reach $\hat{F}$ sooner. The final interval will have unchanged cumulative oil flow. Hence, cumulative oil flow will be strictly smaller. By similar arguments, if cumulative flow is smaller than S, then we should lower γ_0. One implication is that if R_0 increased exogenously, then the equilibrium γ_0 must fall. This would result in a longer first interval, uniformly higher flow during the second interval, a later termination of the second interval, and an unchanged flow during the final interval.

Throughout this section, we have assumed that marginal utility is bounded ($U'(0) < \infty$) and that the rate of drilling is constrained ($\bar{a} < \infty$). If marginal utility were unbounded at the origin, then the equilibrium would be similar to what we have just described but without the final interval: drilling would never cease completely. This case differs from that in section 4 because the drilling rate is constrained. If instead there were no constraint on the rate of drilling, then the first phase would disappear. At the first instant, the planner would drill wells at an *infinite* rate (a "pulse") so that equation (19) would hold immediately at $t = 0$ and would continue to hold until drilling ceased.