Supplemental Appendix

The Dynamics of Evasion: The Price Cap on Russian Oil Exports and the Amassing of the Shadow Fleet

Diego S. Cardoso,[a] Stephen W. Salant,[b] & Julien Daubanes[c]

Contents

[a]Department of Agricultural and Consumer Economics, University of Illinois Urbana-Champaign.

[b]Department of Economics, University of Michigan, and Resources for the Future (RFF).

[c]Department of Technology, Management and Economics, Technical University of Denmark, Center for Energy and Environmental Policy Research, Massachusetts Institute of Technology, and CESifo. Corresponding author: jxada@dtu.dk.

A Sensitivity Analysis of the Baseline Model

In this appendix, we perform sensitivity analyses on key model parameters to examine the impact of different values on the model's qualitative findings. In particular, we examine how model outcomes vary under different price elasticities of demand and non-Russian supply, discount rates, and marginal investment costs, and whether those changes affect our findings.

A.1 Price Elasticity of Demand

The price elasticity of world demand for oil (ϵ_D) plays an important role in determining price levels in response to shifts in Russian exports. To assess how this parameter affects model outcomes, we consider price elasticities of demand (ϵ_D) varying in absolute value between 0.05 and 0.75—six times larger than the baseline elasticity of -0.125.[1] The curves in Figure A1 compare the present value of Russian profits under a service ban and a \$60 cap for a wide range of elasticities. In this sensitivity analysis, we recalibrate ϕ in each run so as to reproduce the same procedure used in the baseline to match initial and ninth-period capacity estimates.

Based on Figure A1 we reach two conclusions. First, we note that the difference in profits between the cap and ban policies is affected by the choice of the price elasticity of demand—and may even flip sign. Lowering the absolute value of ϵ_D towards zero rapidly amplifies the difference, which is a consequence of the profits under the ban increasing faster than under the cap. Moving in the opposite direction, increasing the magnitude of the price elasticity can change the sign of the difference in profits: Roughly doubling the baseline elasticity leads to flipping the ranking, with the cap policy delivering slightly larger present value of profits than then service ban.[2] Further inspection shows that these policies result in the same present value of profits at $\epsilon_D \approx 0.23$.

The second key message from Figure A1 is that substantial increases in $|\epsilon_D|$ only slightly increase the differences in profits. We observe that as the magnitude of the elasticity grows even beyond typical empirical estimates, the curves flatten out, so that the gap between profits in each case remains stable. Therefore, even though a more elastic demand would flip the ordering of profits

[1] In the paper, we report that removing all Russian oil from the market would have driven the world price to approximately \$150 ($P(Z_0) \approx 150$) under the baseline elasticity. In comparison, an elasticity of -0.05 or -0.75 would lead this price to \$376 or \$89, respectively.

[2] For reference, empirical estimates of this elasticity typically range between 0 and -0.4, depending on the estimation approach, time horizon, and data used. See Kilian (2022) for a detailed discussion of the estimation of price elasticities in the oil sector.

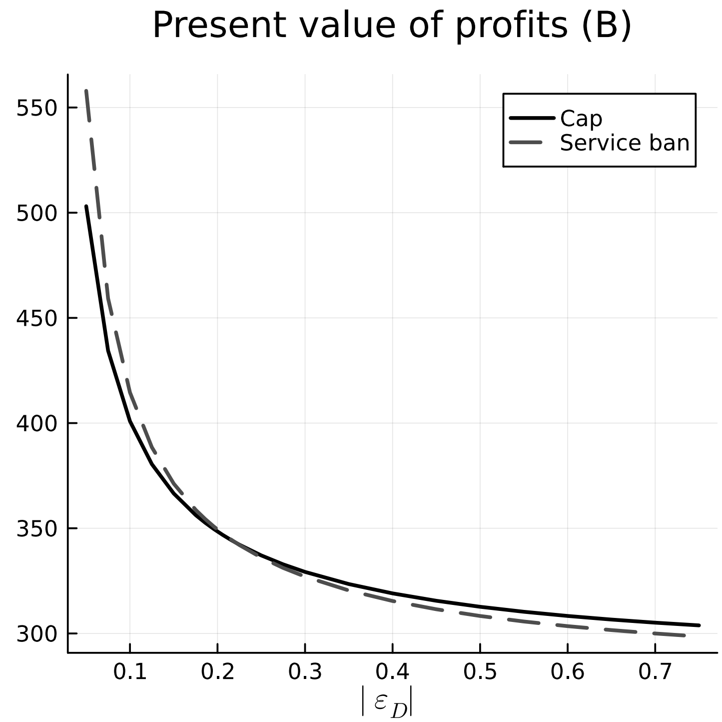

Figure A1: Present value of profits under different policy scenarios for a range of price elasticities of demand. Values are reported in billions of dollars in present value using the baseline discount rate (15% per year) until the sanction termination (80 quarters).

under the service ban or the cap, both policies continue to result in very similar profits in present value.

A.2 Discount Rate

Next, our analysis considers discount rates from 0 to 30% per year—twice as high as the baseline rate of 15% based on Russia's actual cost of capital. A very high discount rate, such as the one used for the upper bound of this analysis, can be interpreted as short-termism or myopia of Russia's decision making, implying that it would be placing an excessive weight on the first few quarters relative to the long run.

Figure A2 shows how the difference between the present value of profits between the cap and the ban policies varies with different combinations of the discount rate and elasticity. Here, we also recalibrate ϕ in each run to enforce the same calibration procedures used in the baseline. Negative values indicate that the PV of profits is lower under the baseline cap ($60) relative to a ban policy.[3]

[3] The simulations reported in Figure A2 use different discount rates for Russia. However, the displayed present

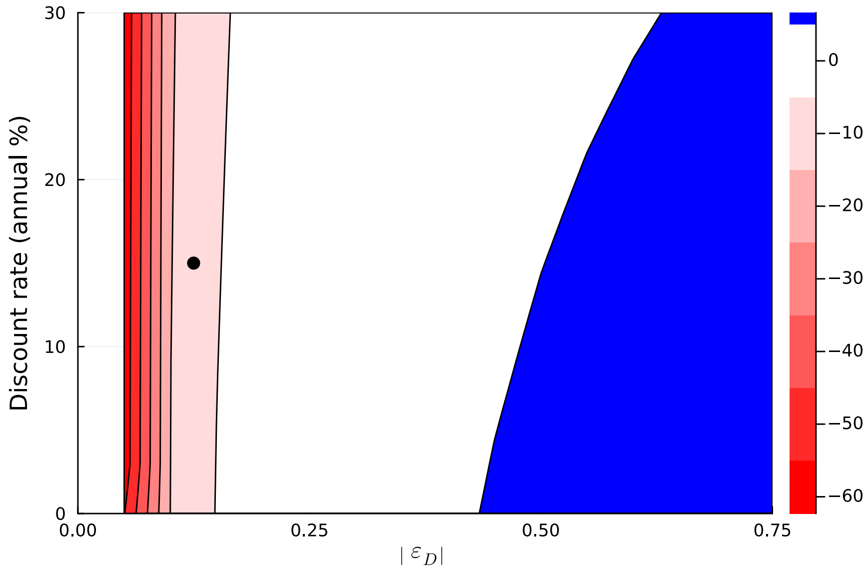

Figure A2: Difference in profits between the cap and the ban policies for various combinations of the discount rate and the price elasticity of demand. The black dot indicates the combination of parameters used in the baseline model. Values are reported in billions of dollars in present value discounted at the baseline discount rate (15% per year) until the sanction termination (80 quarters).

The almost vertical level curves in Figure A2 demonstrate that the ordering in the outcomes between the cap and ban policies is largely insensitive to the choice of the discount rate. Moreover, we highlight that for a wide range of combinations between discount rates and elasticities of demand, the absolute difference in the profits under both policies is below 5 billion dollars (the area in white). This result is primarily due to the calibration of the marginal investment cost function F'. Since we fit the slope of this function to match observed data, an increase in the discount rate leads to a proportionally lower marginal investment cost so as to match the observed level of investment.

To further illustrate how the equilibrium paths under competition are fairly insensitive to discount rate choices, Figure A3 shows a panel of trajectories for the cases of halving and doubling the baseline rate. These panels show that trajectories under the cap policy vary little, and those under the service ban are essentially the same. Therefore, our qualitative baseline results remain robust within a reasonable range of discount rates.

value of profits is calculated using a common rate to ensure fair numerical comparisons across different scenarios.

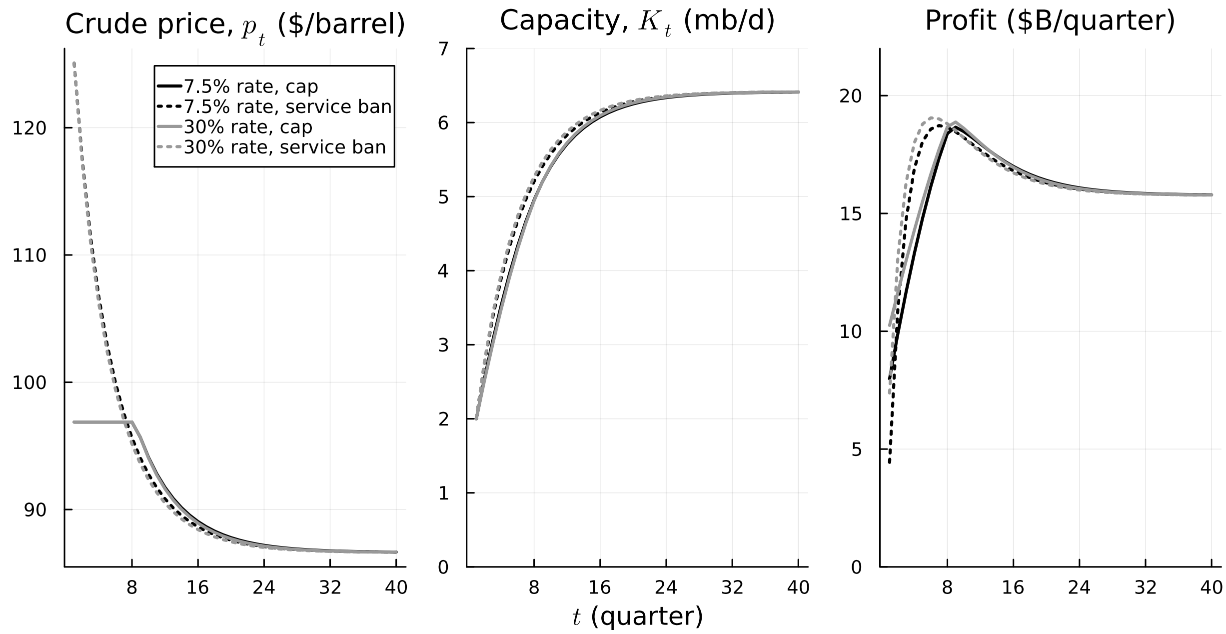

Figure A3: A comparison of the trajectories under a price cap sanction (solid) and a service ban (dashed). Black lines represent trajectories simulated using a 7.5% annual discount rate, whereas gray lines represent these trajectories using a 30% discount rate. All scenarios are based on the baseline price elasticity of demand ($\epsilon_D = -0.125$). These panels display only the first 40 quarters of the 80-quarter simulation.

A.3 Marginal Investment Cost

Next, we consider the impact of changing the parameter ϕ that determines the marginal cost of expanding the shadow fleet. In our simulations, we calibrate ϕ so that the solution of the baseline model (with a $60 cap) replicates the observed expansion of exports to non-Western countries. In this analysis, we vary ϕ to assess its effect on the differences in profits between the cap and service ban policies.

In Figure A4, we examine how different values of ϕ affect simulation outcomes. The medium ϕ scenario reproduces the baseline simulation, with a calibrated $\phi = 4.102$. The other two scenarios, low and high ϕ adopt respectively half and double the calibrated ϕ. These panels show that a higher marginal investment cost leads to a slower expansion of the shadow fleet in either sanction. Therefore, a higher ϕ corresponds to a longer price plateau and a later peak in profits. However, as the bottom right panel shows, the present value of profits is higher under the service ban relative to a $60 cap, regardless of the value of ϕ.

Figure A5 examines whether the relative ranking of profits under a service ban or cap flips with

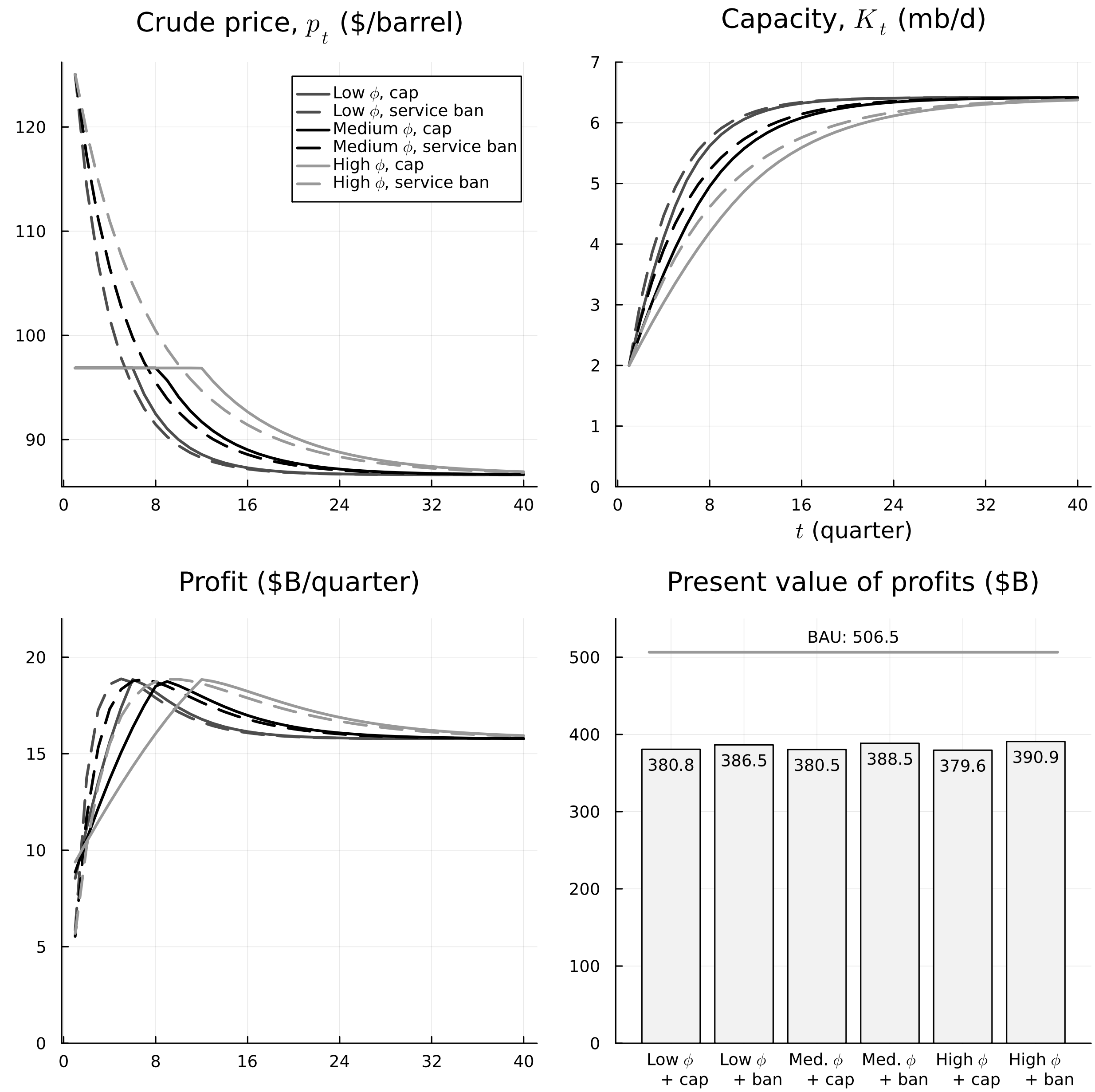

Figure A4: A comparison of prices, capacity, and profits under the service ban (dashed lines) vs. a $60 price cap sanction (solid) with different marginal expansion costs (ϕ). The line graphs display only the first 40 quarters of the 80-quarter simulation.

higher values of ϕ. As this figure indicates, the curves representing profits under each sanction do not cross even for values of ϕ that are ten times larger than the calibrated one. Moreover, we observe that the gap between profits increases: A higher ϕ leads to higher world prices under the service ban, which boosts short-run profits relative to a longer plateau seen under the cap.

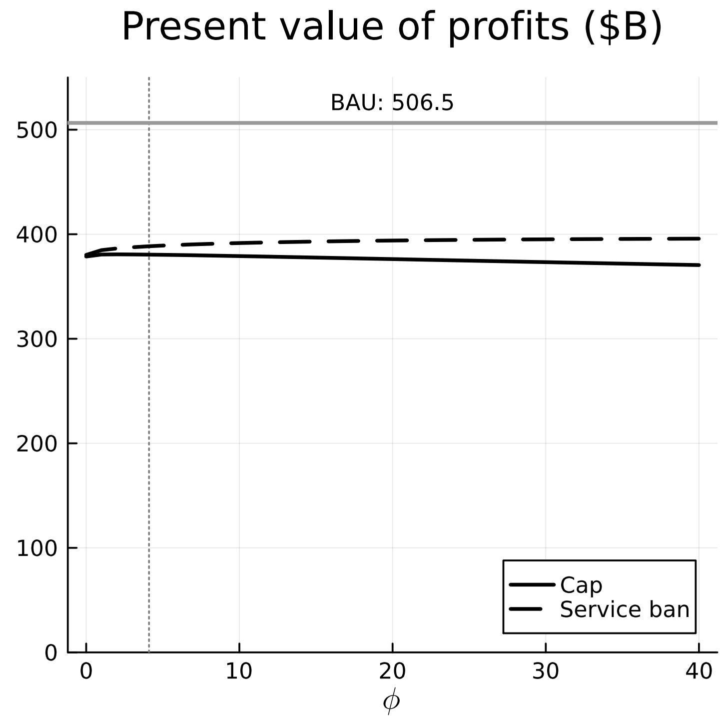

Figure A5: Present value of profits under the service ban or $60 price cap sanction for different levels of the marginal investment cost (ϕ parameter). The vertical dashed line in gray indicates the calibrated ϕ in the baseline model.

Discussion of Alternative Functional Forms for the Investment Cost. Besides changing the slope of the marginal investment cost function F', one could also consider two additional changes. First, we could introduce a positive intercept so that $F'(0) > 0$. The practical result of a positive intercept would be that the size of the shadow would converge to a lower value, as investment levels would stop short of the point where $C'(R^*) = P(Z_0 + R^*) - d$. Second, we could make F a function of K_t to introduce dynamic convexity in investment costs. Again, this modification would result in the shadow fleet converging to a value lower than R^*. Although such modifications are technically possible, we opt for a parsimonious definition of F for two reasons: (i) The shadow fleet expansion is still in progress as we write this paper, so we do not have reliable data to calibrate parameters governing the convergence point, and (ii) our adopted F allows us transparently to link the final size of the shadow fleet with calibrated marginal production costs and producer prices, as represented in the diagram in Figure 1.

B Relaxing Baseline Assumptions

B.1 Endogenous Non-Russian Supply

In the baseline version of our model, we have assumed that non-Russian supply of oil is constant (at Z_0). If instead oil supply from the rest of the world was sensitive to the world oil price ($Z_0'(p) > 0$), we find that the price spike caused by the service ban is smaller and, as a result, the service ban might become more harmful to Russia than the cap although both sanctions deliver quantitatively similar profits in present value.

In this section, we examine how a non-zero price elasticity of non-Russian supply affects the outcomes of the cap and service ban. To do so, we maintain equations (1)–(4) but rewrite the market-clearing conditions as

$$p_t = P(Z_t + Q_t + X_t) \tag{B1}$$

$$Z_t = Z_0 \left(\frac{p_t}{p_0}\right)^{\epsilon_Z}, \tag{B2}$$

where $\epsilon_Z \geq 0$ is the assumed price elasticity and p_0 and Z_0 are defined in the baseline model (see Table 1). Thus, our baseline model is equivalent to setting $\epsilon_Z = 0$..

We simulate the extended model adding equations (B1–B2) for two positive values of ϵ_Z: 0.1 and 0.2. For reference, recent estimates of the price elasticity of global oil supply range from 0.01 (Kilian and Murphy, 2014) to 0.15 (Baumeister and Hamilton, 2019). As in the other extensions, we recalibrate ϕ (the slope of the marginal investment cost function) so as to generate capacity expansion trajectories comparable to observation and the baseline model.

Figure B1 displays the outcomes of these simulations. As shown in the top left panel, the more intense expansion of non-Russian supply with a higher ϵ_Z dampens the price spike under the service ban, as well as the level of the initial price plateau under the price cap. Moreover, with additional non-Russian supply, steady state prices are also slightly lowered relative to a fixed Z scenario. Following a lower price spike, incentives for capacity expansion are reduced, and the top right panel shows that a higher ϵ_Z induces both a lower steady state world price.

The impacts on Russian profits are displayed in the lower panels of Figure B1. The bottom left panel shows that as the price responsiveness of non-Russian supply increases, Russian profit

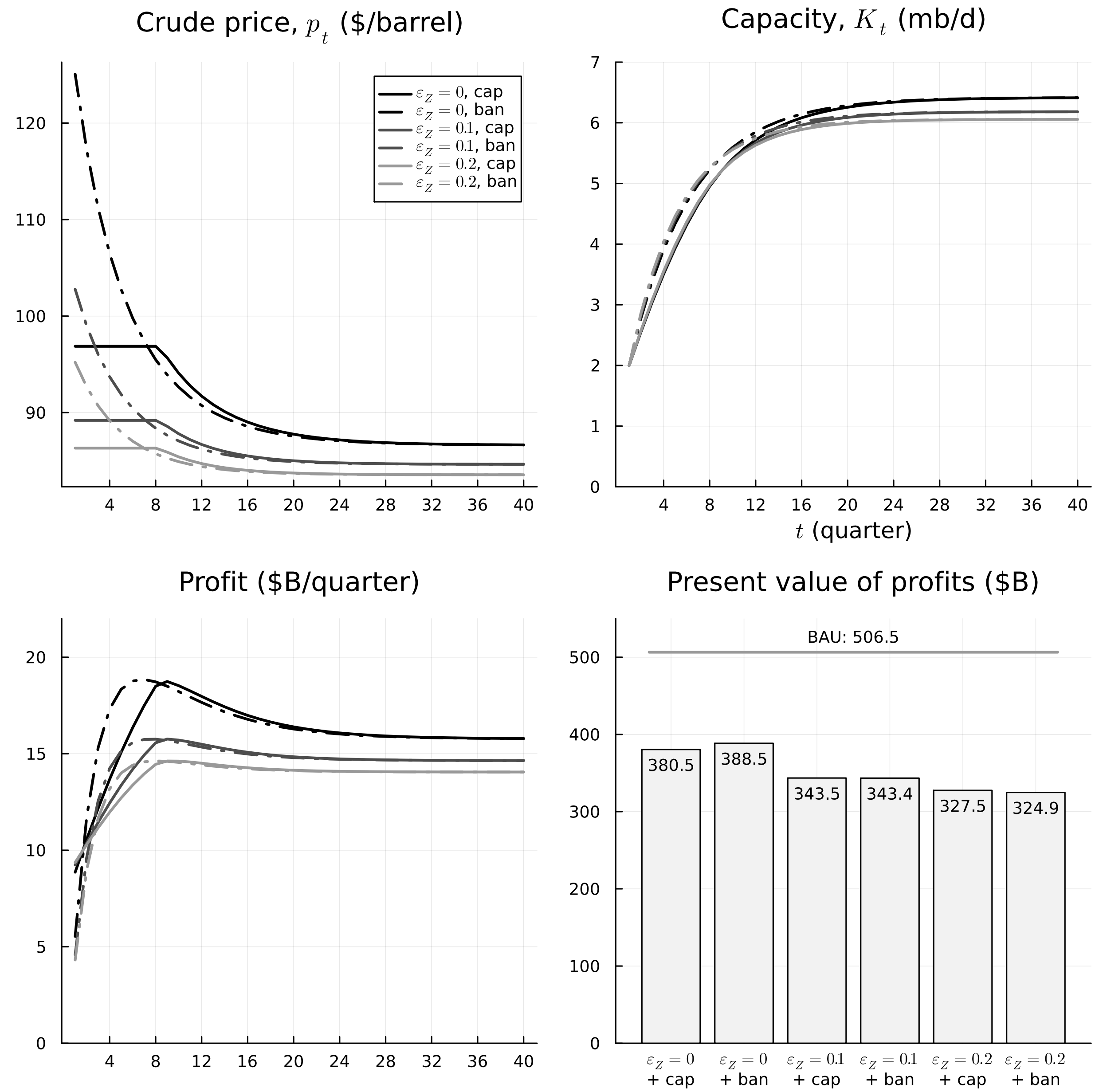

Figure B1: A comparison of the trajectories for different price elasticities of non-Russian export supply (ϵ_Z) when the sanction is a high price cap ($60) or a service ban. Each panel displays quarters 1–40 of the 80-quarter simulation.

opportunities shrink in the early periods, especially in the first eight quarters under the service ban. As a result, the ranking of PV of profits reverses for $\epsilon_Z = 0.1$ or 0.2. However, the present values under both policies remain close.

Next, we simulate both policy scenarios for intermediate values within the considered range of ϵ_Z. As Figure B2 reflects, the cap harms Russia's present value of profits more than the ban at

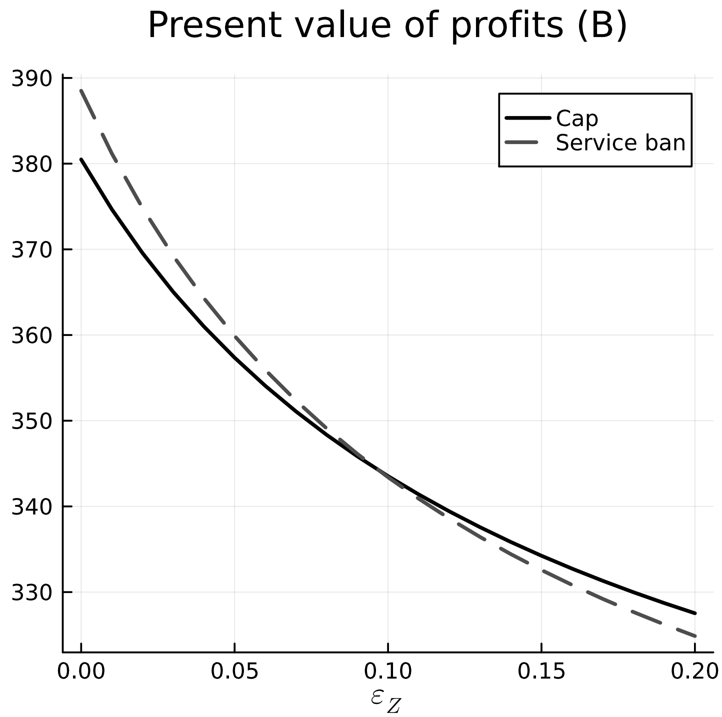

Figure B2: Present value of profits under the service ban and cap for a range of price elasticities of non-Russian supply ($\epsilon_Z \in [0.0, 0.2]$).

$\epsilon_Z = 0$. At the other extreme, the ban is more harmful. This graph also shows that the two policies are equally harmful at $\epsilon_Z \approx 0.10$.

B.2 Curvature of the Marginal Production Costs

Our baseline model assumes affine marginal costs. Nevertheless, empirical estimates of break-even prices of Russian oil assets indicate that the marginal cost curve may have significantly more curvature (Wachtmeister, Gars, and Spiro, 2023). In this Appendix, therefore, we generalize the marginal cost function to allow for varying degrees of curvature.

Previously, we calibrated the affine marginal cost function so it passed through two points. To allow for a more flexible functional form, we define

$$\tilde{C}'(X_t + Q_t) = c_0 + (p_0 - c_0)\left(\frac{X_t + Q_t}{Q_0}\right)^{\chi}, \tag{B3}$$

where c_0, p_0, and Q_0 are defined in the baseline model (see Table 1), and χ measures the curvature of this function passing through the same two points. $\chi = 1$ represents a linear function, whereas higher values of χ bend the curve towards a "hockey stick" form. Figure B3 depicts the marginal cost curve for various degrees of curvature. Note that the strictly convex marginal cost is below the linear marginal cost for aggregate exports below the initial level Q_0 (which is also the maximum in our simulations) of 7.4 mb/d.

We simulate the cap and the service ban policies with more convex marginal costs determined by values of χ at 2 and 8. Here, we also recalibrate the marginal investment cost function (parameter ϕ) in each of the cases to ensure that expansion trajectories under the cap are comparable with the baseline and across scenarios. The results are displayed in Figure B4. We note that a larger χ means that the marginal cost drops further for a given reduction in exports. Under competitive behavior, this means that supply is more inelastic to price changes. Therefore, Russian supply is above the baseline ($\chi = 1$) in all periods under the cap or after the first period under the service ban. The top panel in Figure B4 reflects these mechanisms. The top left shows that the world price drops more quickly under the ban and has longer plateaus under the cap, with both price paths converging to a lower steady state than in the baseline simulations. Similarly, the top right panel shows that a higher supply also means faster expansion under the ban that converges to a higher steady-state capacity, reflecting the lower marginal costs nearing Q_0.

The bottom panels of Figure B4 illustrate how profits change under different values of χ. Profits are generally higher, primarily because costs are lower. Since each scenario is simulated

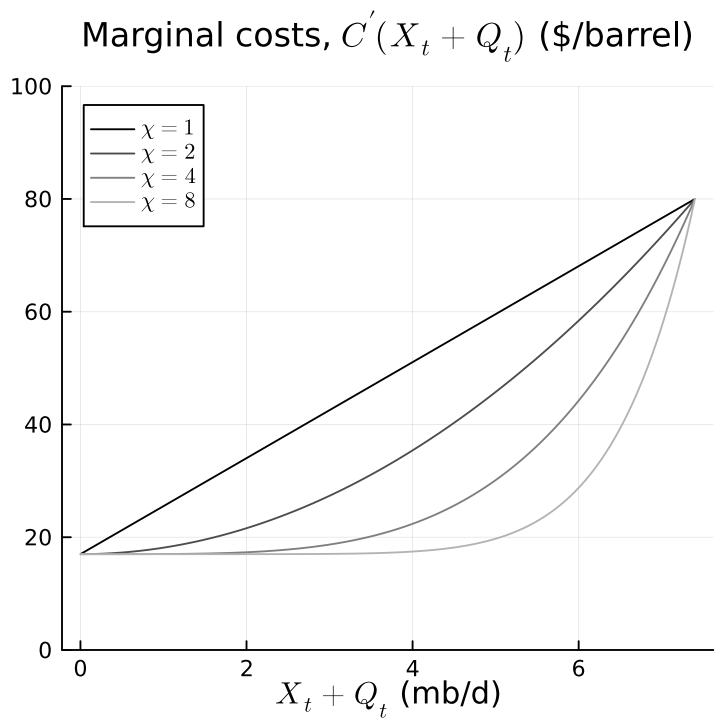

Figure B3: Varying curvature of marginal production costs with different levels of parameter χ.

under different costs, it is difficult to compare the present value of profits across scenarios. For this reason, in the bottom right panel we also plot what BAU profits would be under each χ. The vertical bars show that a higher convexity of marginal costs might lead to a higher present value of profits under the cap relative to the ban. This result is further illustrated in Figure B5, which shows that the ranking flips for $\chi > 2.36$. Nevertheless, both Figures B4 and B5 show that the present value of profits under each policy stays (i) close to each other and (ii) substantially lower than BAU, regardless of χ.

The main implication of χ is the fact that it modulates the supply responses to price changes. As Figure B3 shows, a higher χ leads to a steeper marginal cost curve near the initial export level Q_0. Although our stylized model is not equipped to capture and accurately predict the intricacies of global oil markets—much less so during the turbulent period following the Russian invasion—it is useful to at least gauge key generated outcomes with those observed in data. In particular, we examine average quarterly prices and supply as published in IEA's monthly Oil Market Reports (International Energy Agency, 2022, 2024b) in comparison with model outcomes.

Table B1 reports prices and quantities over different periods and the supply elasticity implied by their changes. In the observed data, we examine Urals prices and total Russian supply in the quarter immediately before the invasion (2021 Q4) and two years later (2023 Q4). This data shows that prices dropped by 9.6% between the end points of this comparison, whereas total Russian supply—including domestic and non-maritime exports—fell by about 5%. The ratio between these changes roughly implies an elasticity of 0.42. This magnitude is substantially above typical estimates of global oil supply (discussed in Appendix B.1), highlighting the exceptionalism of wartime dynamics and the restrictions imposed by the multiple sanctions on the Russian economy.

For the model results reported in Table B1, the initial period reflects data-driven assumptions: slightly rounding up Urals prices from 2021 Q4 and export volumes as reported in International Energy Agency (2024a). The final period is represented by the steady-state solution of the model in each case. Final prices in those cases represent producer prices, which are calculated as the world oil price minus the discount on shadow fleet sales. Here, we highlight how the cases of $\chi = 1$ (baseline) and $\chi = 2$ generate similar prices similar to those in the observed data, thus facilitating the comparisons. Since the model only simulates exports, we adjust implied elasticity calculations by multiplying back the export share of total supply volume. The rightmost column shows that implied elasticities for $\chi = 1$ indicate that this parameter choice generates stronger supply responses than those observed in the IEA data, whereas $\chi = 2$ best approximates those responses among the examined cases. Nevertheless, as indicated in our discussion above, qualitative results are not overturned by increasing $\chi = 2$, for which reason we maintain the parsimonious choice of a simpler functional form in the baseline.

	Initial price	Initial quantity	Final price	Final quantity	Implied elasticity
Observed					
IEA data	78.36	9.90	70.83	9.50	0.42
Model results					
$\chi = 1$	80.00	7.40	71.63	6.42	0.95
$\chi = 2$	80.00	7.40	69.34	6.75	0.50
$\chi = 4$	80.00	7.40	67.59	7.01	0.26
$\chi = 8$	80.00	7.40	66.44	7.18	0.13

Table B1: A comparison of implied elasticities. *Observed* data reports quarterly average Urals prices in 2021 Q4 (initial) and 2023 Q4 (final) in dollars per barrel, and total supply quantities in mb/d for the same periods. *Model results* rows present outcomes under different χ values. For model results, the initial period is $t = 0$, before any sanctions, and the final period is the steady state solution at $t = T$; price columns report producer price (world price minus the shadow fleet discount) and quantities represent exports in mb/d. The implied elasticity is the ratio of percent changes in quantities and prices, adjusted for the export share of total supply ($\approx 74\%$) in the model results.

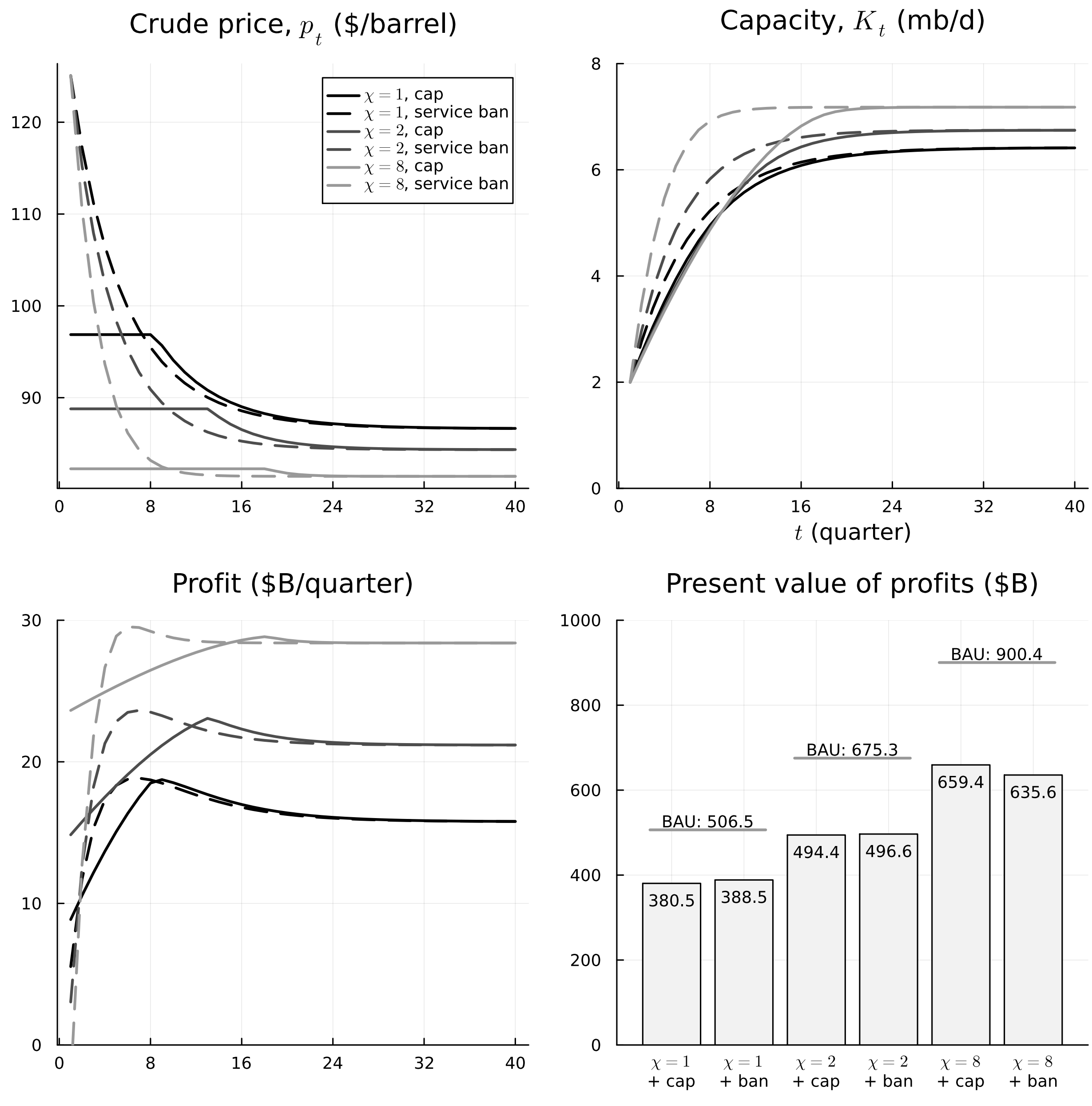

Figure B4: A comparison of the trajectories for different curvatures of the marginal production cost function under a price cap or a service ban policy. $\chi = 1$ defines a linear marginal cost curve, and higher value increase its convexity. Each panel displays quarters 1–40 of the 80-quarter simulation.

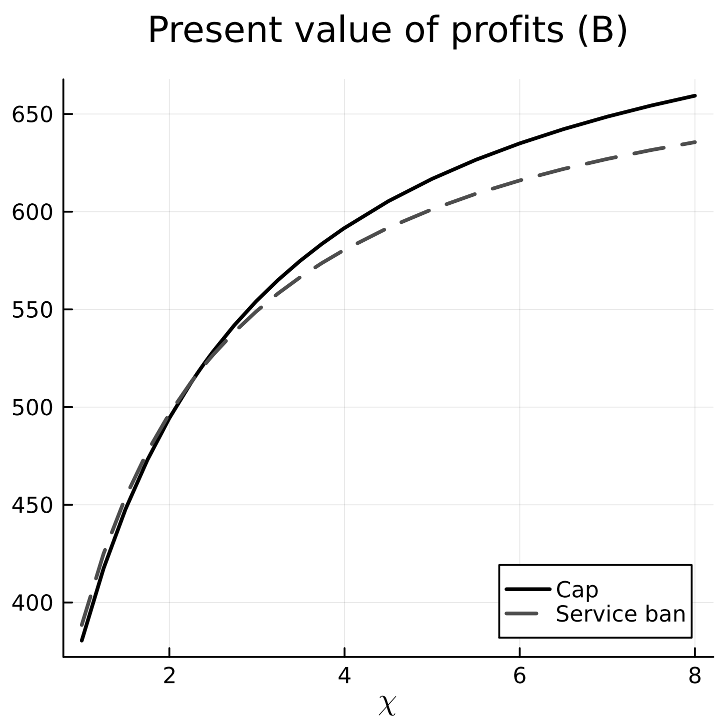

Figure B5: Present value of profits under the service ban and cap for a range of marginal cost curvatures ($\chi \in [1, 8]$).

B.3 Endogenous Discount on Shadow Fleet Sales

In our baseline simulations, the discount at which Russia sells its exports via the shadow fleet is held fixed at $d = 15$, reflecting the differential between Brent and Urals reference prices when it stabilized approximately one year after the actual start of sanctions. In addition to an insurance premium, the literature suggests this discount could also reflect a combination of factors, including higher shipment costs and monopsony power.[4] Kilian et al. (2024) estimate the increased cost of shipping to more distant buyers, finding higher costs notably to China and India. A higher discount may also indicate the increased cost of insurance against being found in violation of the cap.

Johnson et al. (2023) suggest that this discount partly results from the bargaining power of non-coalition importers of Russian oil, observing that the price cap "likely provides negotiating power to oil importers that continue to buy Russian oil above the cap without using [Western] services." Kilian et al. (2024) substantiate that the bargaining positions accounted for observed high initial discounts on Russian oil. Arguably, the opportunity cost of Russia having to sell at the capped price if negotiations with importing countries broke down might account for this bargaining power.

Suppose we assume that the discount is an increasing function of the difference between the world price and the cap. Then if the discount reflects bargaining, the lower the cap, the larger the discount from bargaining. In particular, in the first phase, a reduction in the cap would cause Russia to reduce its sales at the cap, driving up the world price and hence driving up d. In the second phase, nothing is sold at the cap, and tightening it has no effect. Suppose instead the discount reflects the premium on insurance in the imperfect enforcement case. If the audit probability a were increased, the expected price per barrel would fall, less cheating would occur, and the world price would rise. Hence, the discount Russia receives "even on the non-shadow fleet" would rise "due to the risk of being found in violation of the cap."[5]

This subsection extends our baseline model to include an endogenous discount on shadow fleet sales. In this extension, the maximum value of evasion is defined by the arbitrage opportunity,

[4]Buyers in India and China, purchase the limited quantities of Russian oil available to them at $p_t - d$ but must pay the world price p_t to satisfy their remaining demand. Hence, whether the discount d reflects shipping costs, insurance premia, or bargaining, buyers would pay the marginal price p_t and Russia would receive $p_t - d$.

[5]The quotations are from an extremely perceptive referee, whom we thank for pointing out this implication of our model.

which is the difference between the world oil price and either the price cap (for cap scenarios) or the marginal production cost (for ban scenarios). We posit that d represents the portion of the arbitrage value that is absorbed by agents other than Russia (monopsony power from new buyers, transportation, increased insurance costs and risk premia, etc). Based on this, we parametrize an endogenous discount as a function of the arbitrage opportunity offered by evasion,

$$d\left(P(Z_0+K_t)-\hat{p},K_t\right)=\nu_0+\nu\left[P(Z_0+K_t)-\hat{p}\right], \tag{B4}$$

where ν_0 and ν are the parameters to be determined. We highlight that, in equilibrium, K_t itself is a function of $\hat{p}$.

Following our choice of $d=15$ based on data, we interpret this value as the steady state discount, which is obtained when $Q_t=0$ and $X_t=R^*$. We note that Kilian et al. (2024) estimated added transportation costs using the shadow fleet were between \$12 and \$15 per barrel. Therefore, our interpretation could be understood as transportation costs remaining as the long-lasting component of the discount over time. Moreover, we observe that in the weeks following the invasion (March and April of 2022), the gap between Urals and Brent prices achieved its peak, at approximately \$36 per barrel; during the same period, the Brent price peaked at around \$123. In our model, the value of evasion is the biggest in the initial period of the service ban scenario, in which world price peaks and shadow fleet exports are limited to K_1. We use these two observations—the peak shortly after the invasion and the steady state level—to establish the following conditions

$$\begin{aligned}\nu_0+\nu\left[P(Z_0+R^*)-\hat{p}\right] &= 15\\ \nu_0+\nu\left[P(Z_0+K_1)-C'(K_1)\right] &= 36.\end{aligned}$$

For reference, in our calibrated baseline model we have $R^*=6.4$ mb/d, $P(Z_0+R^*)=\$86.62$, $P(Z_0+K_1)=\$125.07$, and $C'(K_1)=\$34.03$. Solving for the parameters, we obtain $\nu_0=6.32$ and $\nu=0.33$.

We build on this idea to expand the model and allow for an endogenous discount that varies with the arbitrage opportunity offered by shadow fleet sales. In doing so, we revise equation (2).

Since $p_t - d_t = p_t - [\nu_0 + \nu(p_t - p_0)]$,

$$X_t \geq 0, [(1-\nu)p_t + \nu\hat{p}] - \nu_0 - C'(Q_t + X_t) - \alpha_t \leq 0, \; c.s.. \tag{B5}$$

We simulate the effects of a \$60 cap and a service ban using this extended model. Figure B6 shows the results of these simulations. As in most of our extensions, we recalibrate the marginal investment cost function for the endogenous discount case to generate capacity trajectories under the cap that are comparable with the baseline and across scenarios. Due to this recalibration, the top panel shows that capacity and price trajectories under the cap perfectly overlay, regardless of the discount model (solid lines). Also for the cap policy, the bottom left panel shows that the endogenous discount model leads to a slightly higher discount during the "price plateau" phase, but ultimately converges to the fixed discount level. This difference results in a small decrease in the present value of profits relative to the baseline model with a fixed discount, as shown in the bottom right panel.

Substantial differences in trajectories occur, however, in the service ban scenario. With the absence of a channel to export at the cap, the value of the arbitrage opportunity is amplified. As the bottom left panel in Figure B6 shows, the endogenous discount in the service ban case sits at a higher level than the fixed baseline value of \$15 throughout the entire simulation. Notably, it converges to approximately \$25 per barrel. This wedge lowers the incentive to capacity expansion, resulting in a lower steady-state fleet (top right panel) and a higher steady-state world price (top left panel). Consequently, with fewer exports and lower effective export prices, Russian profits under a service ban with endogenous discount are more negatively impacted relative to the cap scenario. The drop in the present value of profits with endogenous discount is enough to flip the ranking of profits and create a significant gap between both outcomes—and both remain substantially below BAU profits. Further simulations holding $\nu_0 = 6.32$ constant and varying ν show that both policies result in equal harm to Russian profits with a value of $\nu \approx 0.019$.

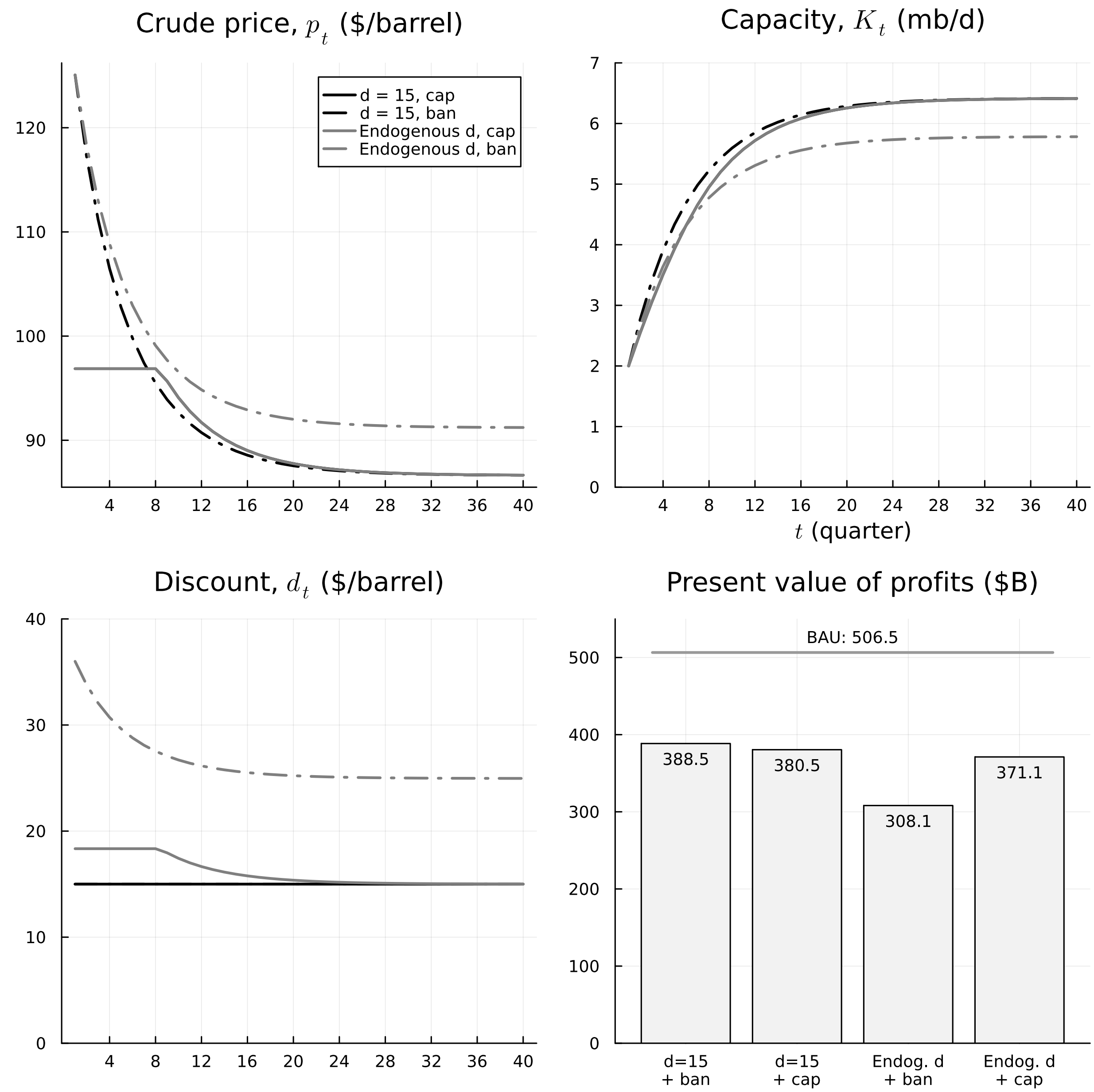

Figure B6: A comparison of the trajectories under a price cap or a service ban policy based on fixed vs. endogenous shadow fleet discount models. Each panel displays quarters 1–40 of the 80-quarter simulation. On the top panels, the curves for "d = 15" and "Endogenous d" perfectly overlap due to the recalibration of the investment cost function.

B.4 Internalization of Price Effects

A paradoxical result in our dynamic model is that Russia, a profit-maximizer, can benefit from a lower price cap even though Russia could have seemingly secured the same benefits without any change in the cap. This result is not an artifact of our price-taking assumption. For example, Turner and Sappington (2024) reach a similar conclusion in their static Cournot duopoly, although, as discussed in our literature review, our results are very different from theirs. Yet, as mentioned in the introduction, such a paradox would not arise if Russia were the only agent fully internalizing its market influence, behaving as a monopolist facing the demand for its oil. In reality, as explained in Section 6, Russia's oil supply behavior has been more consistent with the price-taking hypothesis than with the exercise of monopoly power, partly because of the structural imperfect coordination of Russian oil suppliers. In this appendix, we extend the baseline model to examine cases in which Russia internalizes the intertemporal price effects at various degrees.

We have two objectives: (1) to quantify the present value of Russian profits as a function of the cap for different levels of price effect internalization; and (2) to show the invariance of the present value of Russian profits as the level of internalization approaches the monopoly case.

We modify the problem in Subsection 2.3 by replacing $\{p_t\}_{t=1}^{T}$ with $P(Q_t + X_t + Z_0)$. As a consequence, the first-order conditions (1) and (2) are modified as follows

$$Q_t \geq 0, \qquad \hat{p} + \theta X_t P'(Q_t + X_t + Z_0) - C'(Q_t + X_t) \leq 0, \;\; c.s. \tag{B6}$$

$$X_t \geq 0, \qquad P(Q_t + X_t + Z_0) - d + \theta X_t P'(Q_t + X_t + Z_0) - C'(Q_t + X_t) - \alpha_t \leq 0, \;\; c.s. \tag{B7}$$

The remaining equations, (3) and (4), are unchanged. The term $X_t P'(Q_t+X_t+Z_0)$ in conditions (B6-B7) reflects the fact that a barrel sold at the cap depresses the world price just as much as a barrel sold at the market price. However, in both conditions, parameter $\theta \in [0, 1]$ modulates the degree of internalization of price effects. When $\theta = 0$, the agent behaves as a price taker and does not take into account price responses to changes in supply; when $\theta = 1$, the agent fully internalizes its monopolist position. Values between 0 and 1 represent partial internalization of price effects, which could result, for example, from the lack of internal coordination among Russian producers. In practice, this intermediate case will most likely approximate Russia's aggregate supply decisions, as the hybrid structure of the oil production sector in Russia leads to decentralized decisions and players competing for exports (Henderson and Fattouh, 2016).

We simulate the solutions for $\theta \in \{0, 0.25, 0.5, 0.75, 1\}$ and cap levels between \$30 and \$70 per barrel.[6] This exercise assumes that any price internalization behavior starts at $t = 1$. All simulations use our baseline calibration of the investment cost function F. Unlike other sensitivity analyses and extensions, we do not recalibrate parameter ϕ (the slope of F) because for $\theta \geq 0.5$, there are no values of ϕ that can make the model match observed capacity expansion.[7]

Figure B7 presents the value of profits Russia earns under each internalization degree at different price caps. As expected, the present value of Russian profits is higher under monopoly ($\theta = 1$) than under price taking ($\theta = 0$). For a high degree of internalization ($\theta \geq 0.75$), the present value of profits does not increase as the cap decreases. However, with lower degrees, there is a slight decrease in profits as the cap increases, with the most pronounced cases occurring when θ approaches 0. Despite these differences, we note that a key qualitative finding still stands: The present value of profits under a low cap (30) and a high cap (70) are very similar, regardless of the degree of internalization.

Despite similar results for the present value of profits under either policy, this exercise indicates that changes in θ can affect their ranking. We further investigate this notion in Figure B8, which displays the present value of profits for either policy for $\theta \in [0, 1]$. This plot can be seen as representing how profits vary on a continuum of θ along two vertical slices in Figure B7: one at \$30 and another at \$60. Again, this plot illustrates how close the present values of profits are under each policy, regardless of the level of price effect internalization. Numerical inspection of these simulations shows that both lines cross at approximately $\theta = 0.85$, indicating that profits under a \$60 cap would slightly exceed those under a service ban only under high degrees of internalization of price effects.

To further demonstrate how the different degrees of internalization impact profits, Figure B9 displays the trajectories of prices, capacity, and marginal markups under three scenarios: no internalization ($\theta = 0$), partial internalization ($\theta = 0.5$), and full internalization ($\theta = 1$). Based on

[6] For cap values above \$71, the shadow fleet is never used in the competitive solution. For caps below \$34.1/barrel, the Western services are never used—just as if there had been a service ban. This lower bound corresponds to $C'(\bar{K}_1) \approx 34.1$, i.e., the marginal cost associated with allocating all production to the initial shadow fleet capacity.

[7] This is because higher degrees of internalization always lead to a steady state capacity below the observed level.

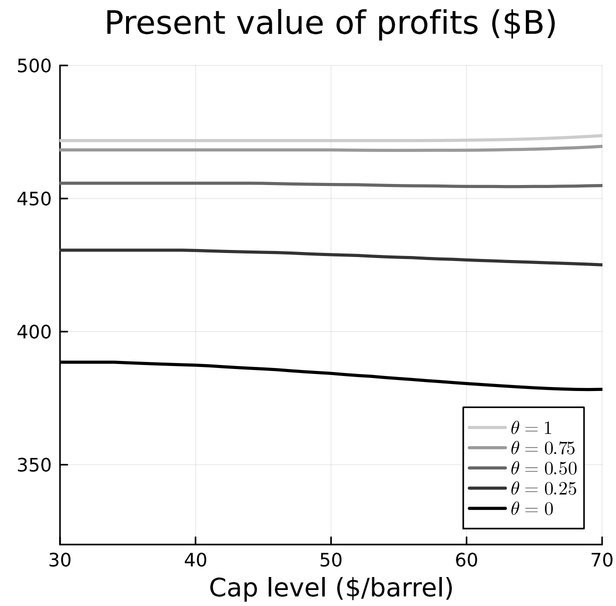

Figure B7: A comparison of profits for different degrees of price effect internalization (θ) under various levels of the price cap. Here, $\theta = 0$ indicates no internalization of price effects, whereas $\theta = 1$ indicates full internalization.

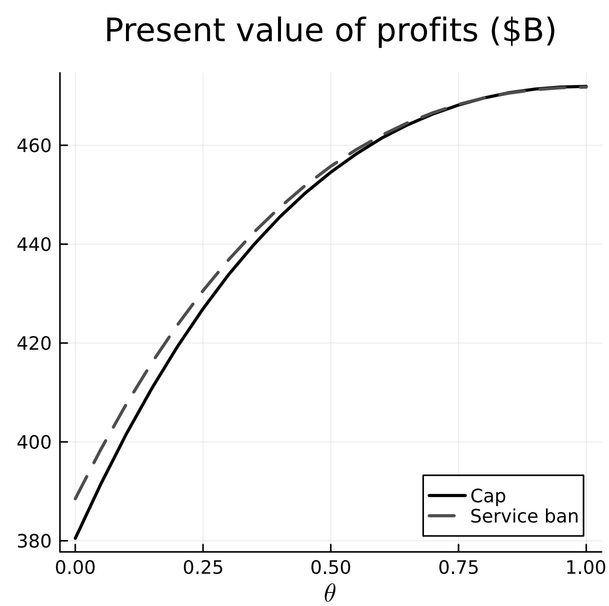

Figure B8: Present value of profits under different policy scenarios for a range of degrees of internalization of price effects. Here, $\theta = 0$ indicates no internalization of price effects, whereas $\theta = 1$ indicates full internalization.

equation B7, marginal markups are defined as the gap between marginal revenue and marginal costs (accounting for the shadow value of a constrained fleet):

$$\mu_t \equiv P(Q_t + X_t + Z_0) - d - C'(Q_t + X_t) - \alpha_t = \theta X_t P'(Q_t + X_t + Z_0). \tag{B8}$$

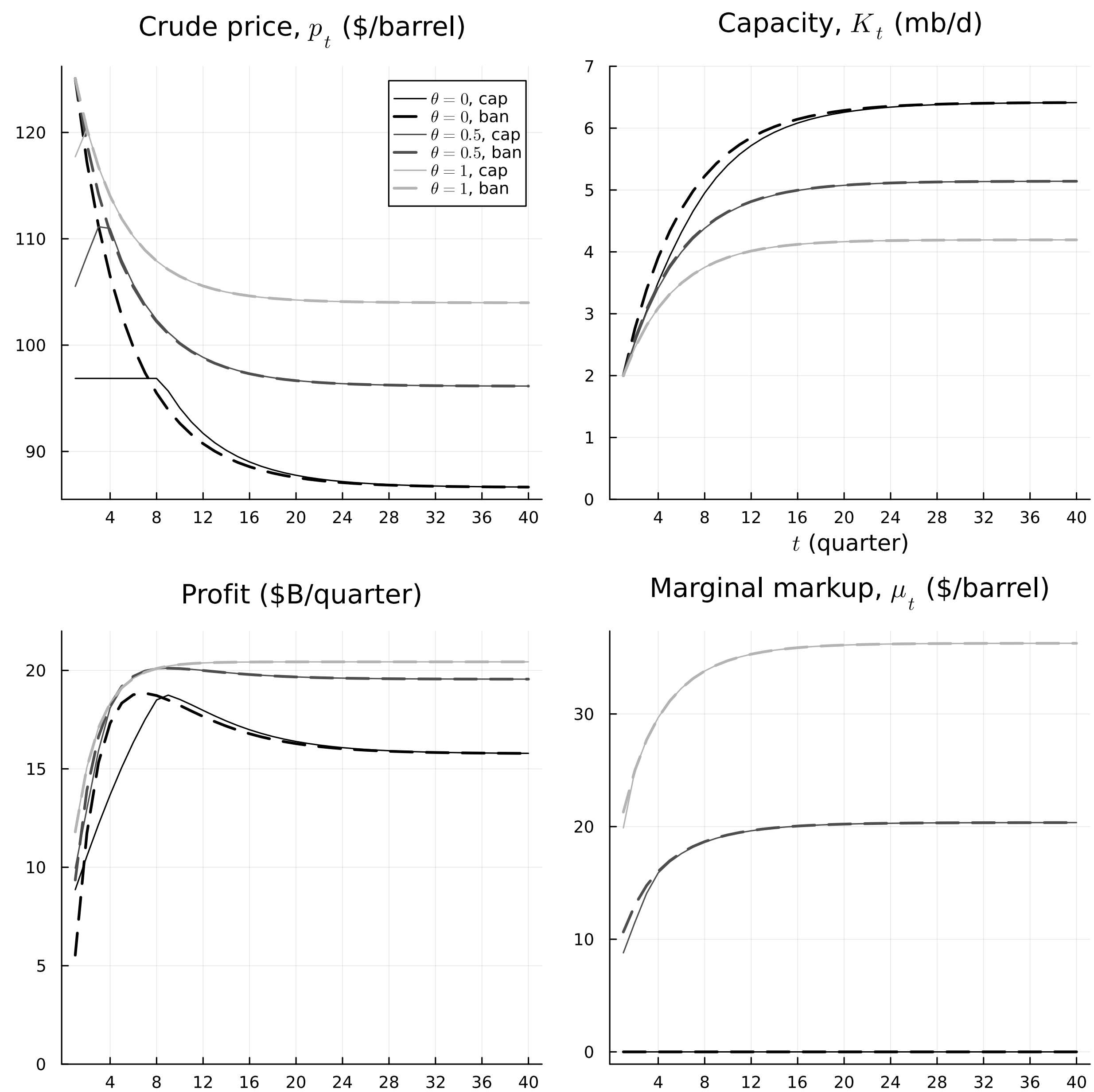

Figure B9: A comparison of the trajectories for different degrees of price effect internalization (θ) when the sanction is a high price cap ($60) or a low price cap (equivalent to a service ban). For $\theta = 1$, the two trajectories in each panel are indistinguishable since they *coincide*. Each panel displays quarters 1–40 of the 80-quarter simulation.

The panels in Figure B9 illustrate the general intuition that a greater internalization of price effects leads to lower supply—in the form of more limited fleet expansion—and sustained higher prices and markups. We note that full internalization trajectories in each of the four panels for the complete service ban and the \$60 cap are *identical* and coincide. The only exception is the price (upper left panel) in the very first quarter. Being unable to enlarge 1^{st}-quarter shadow fleet capacity, Russia would sell a small portion of exports using Western services ($Q_1 \approx 0.7$ mb/d). However, after the 1^{st} quarter, it never uses Western services again, and the price trajectory thereafter coincides with its response to a service ban. For both policies, if Russia fully internalizes its price effects, it expands the shadow fleet far less than a price-taker would. As a result, the world price and the profits are uniformly higher.

While the outcomes for the case of partial internalization generally stand between the extreme cases, the quarterly profits trajectory (bottom left panel) illustrates an important mechanism. While full internalization leads to a monotonic increase in profits, a less-than-perfect internalization causes profits to peak. Further inspection shows that profits peak precisely when capacity reaches the steady-state level of $\theta = 1$. However, due to an underestimation of price effects under partial internalization, the agent continues to expand capacity, thus lowering quarterly profit levels.

B.5 Delayed Implementation and Anticipatory Behavior

Our baseline model assumes the immediate implementation of sanctions, with the shadow fleet expansion occurring while sanctions are in effect. The baseline parametrization of an initial evasion capacity $\bar{k} = 2$ mb/d is based on pre-invasion levels and reflects the impacts of policies implemented before Russia expanded its shadow fleet. However, in reality, the \$60 cap on Russian crude oil came into effect on December 5, 2022—a little over nine months after the invasion of Ukraine. By the time this sanction was implemented, Russia was already amassing its shadow fleet and diverting exports, most notably to India, China, and Turkey. At the end of 2022, Russian average exports to the "non-coalition" regions (China, India, Turkey, and the Middle East) had increased by 1.4 mb/d (International Energy Agency, 2024a). This anticipatory behavior gave Russia a head start, investing in evasion capacity and lowering the effectiveness of policies the West implemented later.

In this subsection, we extend the baseline model to examine the impact of delaying the implementation of sanctions on Russian profits. To do so, we assume that no sanctions are implemented in the first three quarters. Then, a service ban or a \$60 cap becomes effective in the 4^{th} quarter. We assume Russia anticipates this schedule and, taking the equilibrium price path as given, optimizes. Figure B10 displays the results of these simulations and compares the outcomes of immediate vs. delayed implementations.

We focus first on the comparison between delayed and immediate implementation of the cap (solid lines). The top left panel in Figure B10 shows that, with a delayed implementation, the world price stays at $p_0 = 80$ for three quarters. The price then jumps up to the plateau and stays at that level until the 8^{th} quarter. Oil prices begin to decline in the 9^{th} quarter for both the delayed and immediate cap implementation scenarios, indicating that cap sales end at that point, regardless of whether there is a three-quarter delay in implementation or not. The decline in prices, however, is slightly slower under a delayed implementation of the cap. This leads to tiny differences in price trajectories, with the biggest price gap of \$0.54 happening in the ninth period. These differences are due to a slightly slower capacity expansion when the implementation is delayed.

To further investigate these small differences in capacity additions, we report the values for the first six quarters in Table B2. Under a delayed implementation of the cap, the returns of an early expansion accrue later. Time discounting then leads to lower investment in the first quarters, with investment growing at the rate of $1/\beta$ ($\approx$4.15%) per quarter in the first three quarters. This mechanism leads to small differences in capacity trajectories, as shown in the top right panel of Figure B10, where the solid curves almost overlay. The cumulative investment in the first three periods is smaller under a delayed cap (1.38 mb/d added) compared to an immediate cap (1.51 mb/d). We note that this simulated expansion is closely aligned with the expansion of 1.4 mb/d in Russian exports to "non-coalition" regions observed in the IEA data.

As a result of this anticipatory expansion, when the delayed cap comes into effect in the fourth quarter, Russia has a shadow fleet capacity of $K_4 = 3.38$ mb/d. This increased evasion capacity softens the more substantial harm of sanctions in the early period, as shown in the bottom left panel of Figure B10. Under delayed implementations, Russia has a higher producer surplus in the first three quarters, and the downward trajectory of profits is the result of increasing investment costs in preparation for the sanctions. When the delayed cap becomes effective, the trajectories of profits under each policy remain close to their immediate implementation counterparts. The bottom right panel in Figure B10 shows that the ability to expand capacity in anticipation of sanctions raises the present value of Russian profits under the delayed cap relative to its immediate implementation. However, despite the significant head start to adjust, these gains are modest: about 5% (service ban) higher profits in present value, with a still substantial gap to BAU profits.

Comparing the immediate and delayed implementation of a service ban, we observe that small differences in trajectories (dashed lines) are due to the same incentives that lead to slightly different investments. The top left panel in Figure B10, in a delayed service ban implementation, the world price jumps up in the 4^{th} quarter, staying slightly above its immediate implementation counterpart as they decline until convergence around the 20^{th} quarter. Once again, these small differences reflect a slower capacity expansion under a delayed service ban, as shown in the top right panel. Table B2 complements this graph by showing that, when the ban comes into effect, Russia would have amassed enough capacity to export 3.65 mb/d. This anticipatory expansion avoids the profit loss in the first three quarters that would have occurred with an immediate implementation, as shown in the bottom left panel in Figure B10. As a result, however, the present value of profits under a delayed service ban is only 3.8% higher than under an immediate sanction.

Finally, these simulations assume Russia had perfect foresight of sanctions, whereas, in reality, the form of sanctions to be ultimately implemented was uncertain. Incorporating expectations over uncertain policy stringency would require a substantially more complex model. However, such an addition would be unlikely to change these results significantly because we observe that expansion trajectories are very similar, whether Russia expects the most stringent policy under consideration (service ban) or a less stringent policy (cap).

	Scenario			
Variable	Immediate cap	Delayed cap	Immediate service ban	Delayed service ban
I_1	0.54	0.44	0.76	0.53
I_2	0.50	0.46	0.63	0.55
I_3	0.47	0.48	0.52	0.57
I_4	0.43	0.44	0.43	0.47
I_5	0.38	0.40	0.35	0.39
I_6	0.34	0.35	0.29	0.32
$I_1 + I_2 + I_3$	1.51	1.38	1.91	1.65
K_4	3.51	3.37	3.91	3.65

Table B2: A comparison of the trajectories of investments (capacity additions), in the first six quarters. The two bottom rows show the cumulative investments until the 3^{rd} quarter and the capacity starting in the 4^{th} quarter, given the initial capacity of 2 mb/d. All values are mb/d.

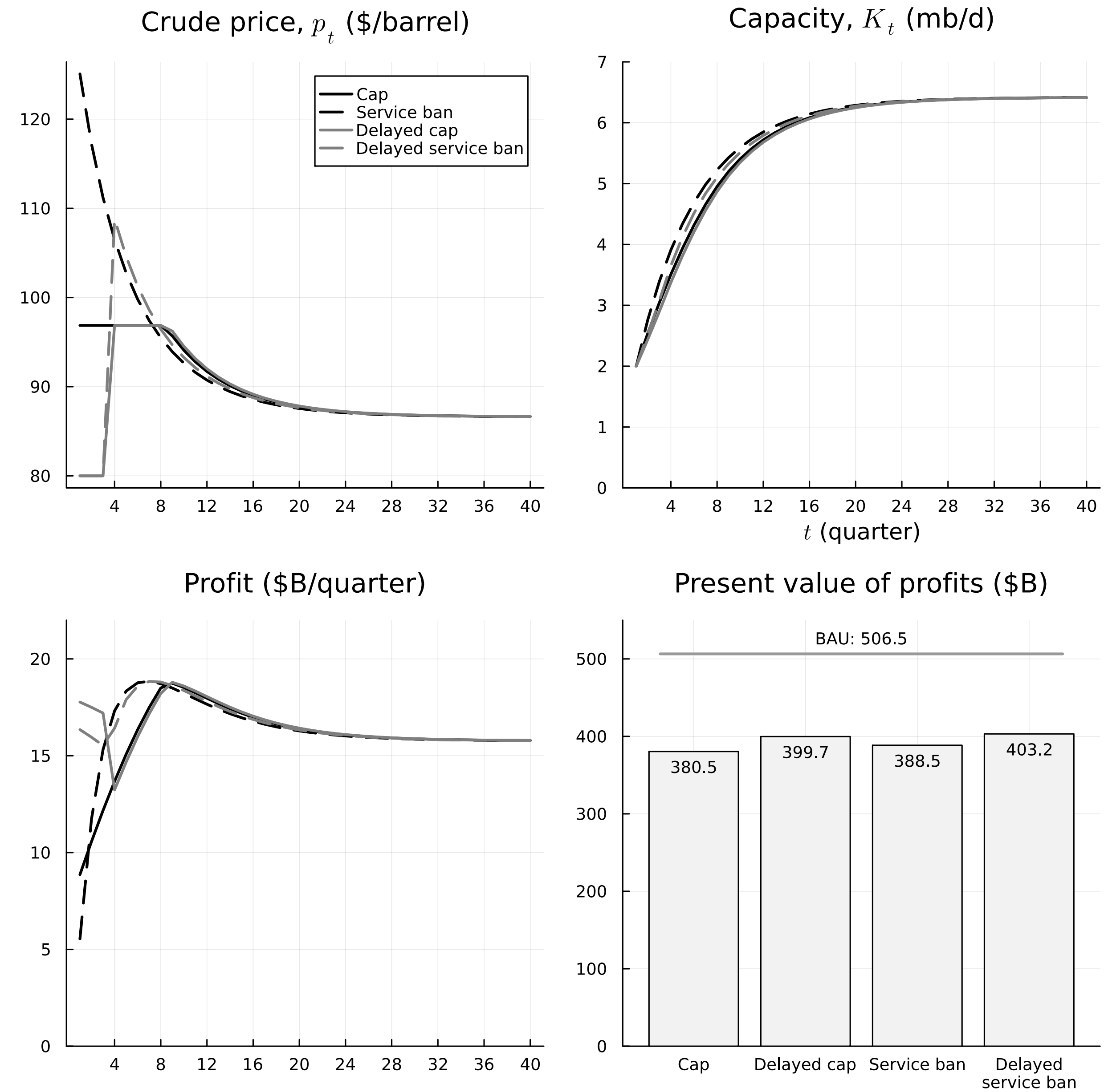

Figure B10: A comparison of the trajectories under a price cap or a service ban policy implemented either immediately or delayed by four quarters. Each panel displays quarters 1–40 of the 80-quarter simulation.

C Targeting the Shadow Fleet while Changing the Enforcement of the Price Cap

The following simulations extend those presented in Subsection 5.3 of the main text, entitled "Unanticipated Targeting of the Shadow Fleet" to the case of sanctions targeting the shadow fleet paired with changes in the enforcement policy.

We maintain the \$60 cap and simulate unanticipated targeting, examining the effect of an unexpected loss in the shadow fleet in the 12^{th} quarter that reduces the size of the shadow fleet to the level Russia had attained 8 quarters earlier: from K_{12} to K_4.[8]

We examine two cases: In the first, auditing occurs only 20% of the time, so cheating occurs prior to targeting. In the second, the cap is perfectly enforced ($a = 1$), so that no cheating occurs prior to targeting.

In each case, the level of enforcement *after* the 12^{th}-quarter targeting can remain the same or can change: Lax enforcement ($a = .2$) prior to the targeting can be replaced by perfect enforcement afterwards, or perfect enforcement prior to the targeting can be relaxed afterwards.[9]

Consider first the case where enforcement is initially lax. The solid lines in Figure C1 show the case where enforcement is lax and there is no targeting. The dashed lines show the case in which the lax enforcement level is maintained after the fleet loss. In comparison to the case without a fleet loss, Russian cheating jumps up to offset the fleet loss, so the world price is unaffected. This illustrates case (i). Targeting in case (i) reduces the post-targeting present value of Russian expected profits (i.e., discounted to $t = 12$). In the bottom right panel of Figure C1, the reduction in the post-targeting present value of expected profits is 1.87%—from \$380.6 to \$373.5 billion.

In response to the targeting, cheating jumps up but subsequently declines monotonically, ending in the 36^{th} quarter instead of the 28^{th} quarter. Until it ends, the world price does not change. The unexpected destruction of the fleet effectively shifts back the trajectory capacity, eventually converging to the steady state level (top right panel).

We can also simulate what happens if targeting and tightening enforcement occur simultane-

[8]Note that Subsection 5.3 considers targeting at K_{12} that reduces capacity to the level attained 4 quarters earlier. Therefore, the targeting reported in this Appendix represents a larger capacity reduction.

[9]With $a = 0.2$ and $\tau = 10$, capacity that had reached $K_{12} \approx 4.2$ mb/d suddenly drops to $K_4 \approx 2.9$—a reduction of 31%; with $a = 1$, capacity that had reached $K_{12} \approx 5.7$ mb/d suddenly drops to $K_4 \approx 3.5$—a reduction of 39%.

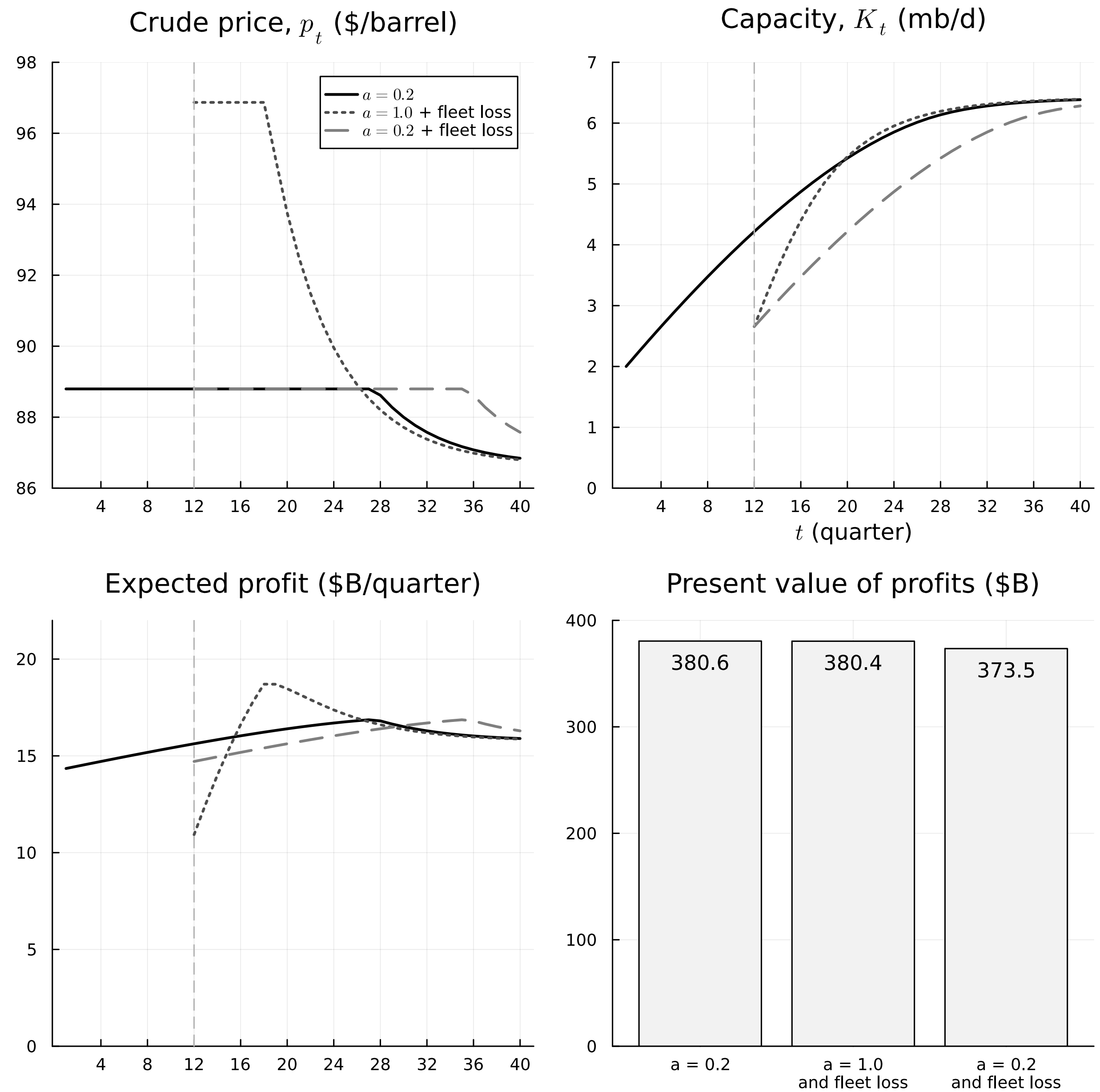

Figure C1: A comparison of the outcomes with and without unexpected changes after the 12^{th} quarter. The solid line represents the low enforcement case ($a = 0.2, \tau = 10$) throughout without shadow fleet loss. The dotted line represents the case with the destruction of the shadow fleet accumulated between the beginning of the 4^{th} and 12^{th} quarters and ramping up to perfect enforcement ($a = 1$) after the beginning of the 12^{th} quarter. The dashed line shows the same shadow fleet destruction but maintaining low enforcement. The present value of profits is calculated for post-targeting periods, discounted to $t = 12$. The line graphs display only the first 40 quarters of the 80-quarter simulation.

ously. Suppose we begin with lax enforcement and then simultaneously target and tighten enforcement. We can break this into two steps. If targeting occurred without tighter enforcement, we have just seen that the present value of Russia's expected profits would fall by a small percentage. If we now tighten enforcement with the fleet capacity held constant, the present value of Russian expected profits increases by a small percentage. As the bottom right panel reflects, the joint effect of targeting and tightening in this case is to lower the Russian post-targeting present value of expected profits imperceptibly (by .05%)—from \$380.6 billion to \$380.4 billion.

The dotted lines in Figure C1 indicate that switching to perfect enforcement would shut down the cheating channel and would cause the world price of oil to jump up (top left panel). Although aggregate exports jump down, all use of Western services is authorized; Russia no longer cheats because of the increased stringency of enforcement. As a result, the lost capacity is rebuilt in only 8 quarters rather than 28 quarters (top right panel). Consequently, with tightened enforcement, Russian profits are higher merely 3 quarters after the unexpected loss of shadow fleet capacity.

Figure C2 shows the results for the case of perfect enforcement of the cap before the 12^{th} quarter. The solid line reproduces the baseline results. The dotted lines represent the trajectories after the 12^{th} quarter targeting with no change in the enforcement level. This illustrates our case (iii). The top left panel reveals that an unexpected reduction in the shadow fleet causes an upward jump in the world oil price. This loss in fleet capacity induces Russia to resume sales under the cap for the next four quarters. The bottom left panel shows that the unexpected fleet loss causes Russian profits to jump down in the short run when compared to the trajectory without capacity loss; however, profits after the fleet loss overtake profits had there been no fleet loss after only four quarters. As the bottom right panel reflects, targeting while maintaining perfect enforcement causes the post-targeting present value of Russian profits to increase by 1.2%.—from \$384.3 to \$389.0 billion.

The dashed lines in Figure C2 illustrate the case where enforcement of the cap is relaxed at the same time that the fleet is targeted. As the top left panel reflects, the world price would jump down. Despite the unexpected lower shadow fleet capacity, equilibrium prices are initially lower due to a more than offsetting upward jump in cheating. Since unauthorized sales earn nearly as much as shadow fleet sales, the fleet is rebuilt only slowly. As the bottom right panel reflects, relaxing enforcement when targeting causes the post-targeting present value of Russian expected profits to fall by 1.7%—from \$384.3 billion to \$377.8 billion.

Suppose auditing is sufficiently frequent that there is no cheating on the \$60 cap. If targeting in

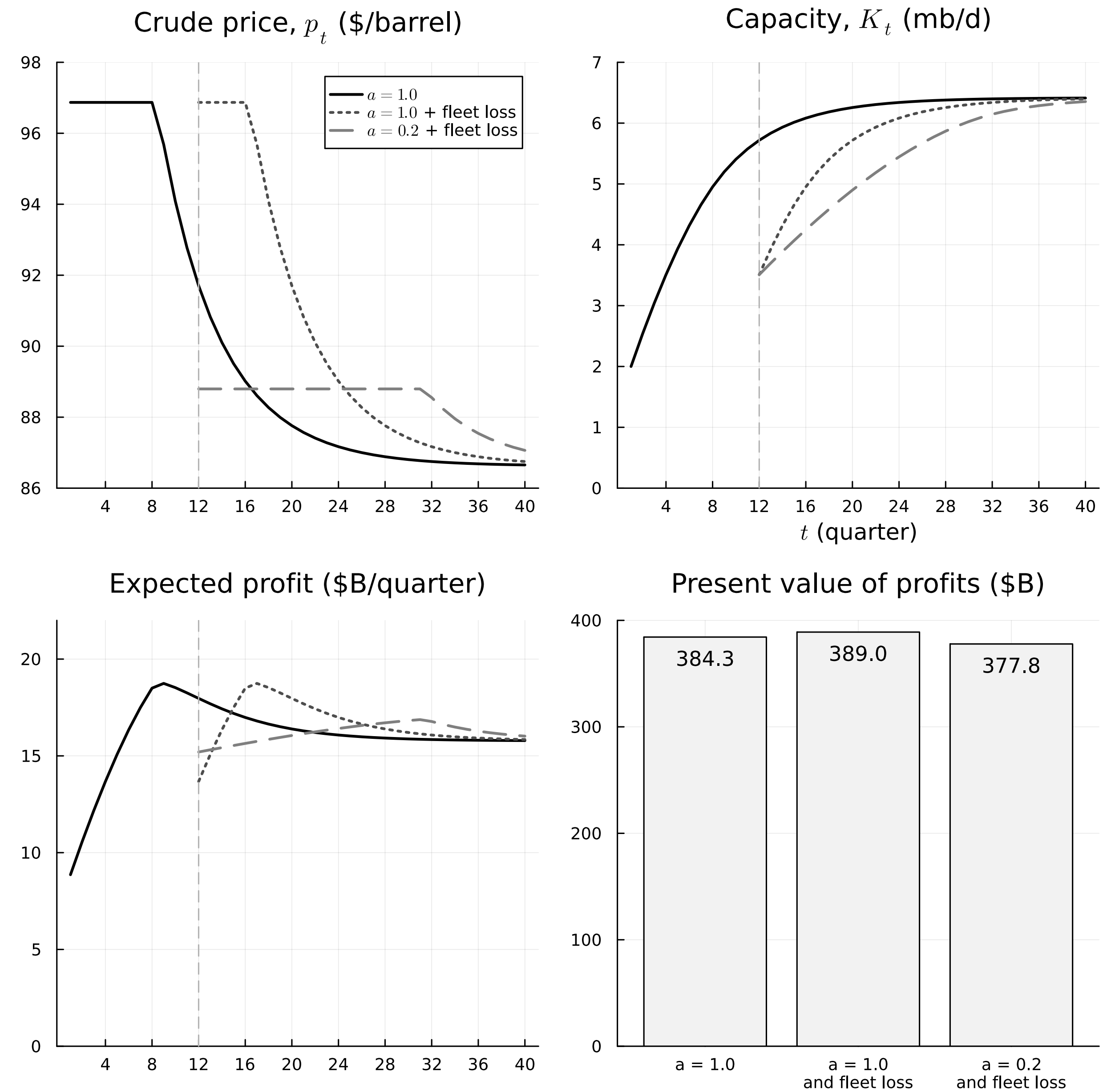

Figure C2: A comparison of the outcomes with and without unexpected changes after the 12^{th} quarter. The solid line represents the perfect enforcement case ($a = 1$) throughout without shadow fleet loss. The dotted line represents the case with the destruction of the shadow fleet accumulated between the beginning of the 4^{th} and 12^{th} quarters while maintaining perfect enforcement. The dashed line shows the same shadow fleet destruction but with low enforcement ($a = 0.2, \tau = 10$) after the beginning of the 12^{th} quarter. The present value of profits is calculated for post-targeting periods, discounted to $t = 12$. The line graphs display only the first 40 quarters of the 80-quarter simulation.

the 12^{th} quarter were *combined* with a policy that prevented the rebuilding of the shadow fleet, then the effect on the present value of Russian profits would depend critically on how much capacity was left after targeting.[10] In the extreme case where fleet capacity was eliminated entirely, the post-targeting present value of Russia's profits over the remaining 69 quarters would drop from \$384.3 billion in the baseline simulation to \$230 billion, a reduction of 40%. In that case, all of Russia's exports would be sold at the ceiling price. Since there would be no Russian sales at the world price, a tighter cap—although it would elevate the world price—would harm Russia even more, consistent with Hotelling's lemma.

[10]We are indebted to Catherine Wolfram for suggesting that we analyze the policy combination of targeting and blocking the rebuilding of the shadow fleet.

References

Baumeister, C., and Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review*, *109*(5), 1873–1910.

Henderson, J., and Fattouh, B. (2016, February). Russia and OPEC: Uneasy partners. *Oxford Institute for Energy Studies*. Retrieved from https://www.oxfordenergy.org/wpcms/wp-content/uploads/2016/02/Russia-and-OPEC-Uneasy-Partners.pdf

International Energy Agency. (2022). *Oil Market Report - January 2022.* Paris. Retrieved from https://www.iea.org/reports/oil-market-report-january-2022

International Energy Agency. (2024a). *Average Russian Oil Exports by Country and Region, 2021-2023.* Retrieved 2024-10-02, from https://www.iea.org/data-and-statistics/charts/average-russian-oil-exports-by-country-and-region-2021-2023

International Energy Agency. (2024b). *Oil Market Report - January 2024.* Paris. Retrieved from https://www.iea.org/reports/oil-market-report-january-2024

Johnson, S., Rachel, L., and Wolfram, C. (2023). Design and Implementation of the Price Cap on Russian Oil Exports. *Journal of Comparative Economics*. doi: 10.1016/j.jce.2023.06.001

Kilian, L. (2022). Understanding the Estimation of Oil Demand and Oil Supply Elasticities. *Energy Economics*, *107*, 105844. doi: 10.1016/j.eneco.2022.105844

Kilian, L., and Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, *29*(3), 454–478.

Kilian, L., Rapson, D., and Schipper, B. (2024). *The Impact of the 2022 Oil Embargo and Price Cap on Russian Oil Prices* [Working Paper]. Retrieved 2024-12-22, from https://sites.google.com/site/lkilian2019/research/working-papers

Turner, D. C., and Sappington, D. E. (2024). On the Design of Price Caps as Sanctions. *International Journal of Industrial Organization*, *97*, 103099.

Wachtmeister, H., Gars, J., and Spiro, D. (2023). *The Price Cap on Russian Oil: A Quantitative Analysis* [SSRN Scholarly Paper]. Retrieved from https://papers.ssrn.com/abstract=4497485 doi: 10.2139/ssrn.4497485