

## Peering Behind the Veil: Trends in Types of Contracts between Private Health Plans and Hospitals

### APPENDIX

#### I. Creating the Analytic Data Set and Validation Data Sets

Appendix Figures 1a and 1b illustrate the sequential application of exclusion criteria to create our analytic and validation data sets.

##### [Analytic Data Set and Colorado Medicaid Fee-for-service \(FFS\) Validation Data Set](#)

The analytic data set, which contains inpatient hospital claims data from private health plans, and the Medicaid FFS validation data set are both extracted from the Colorado All Payer Claims Database (CO APCD), as illustrated in Appendix Figure 1a. We first extract from the CO APCD all hospital inpatient claims covering an entire episode of hospital stay with one service provider. We only keep claims for inpatient services performed at short-term general hospital, as identified by AHA data. We then segment claims into an analytic data set if covered by private health plans, and a validation data set if covered by Medicaid FFS. We exclude claims processed as secondary, covered by Medicare Supplemental plans or if member's age at the time of service was 65 or older. Finally, we exclude claims with allowed amounts equal to zero, as these likely represent bundled or capitated payments; claims with allowed amount greater than charges, as these are likely coding errors; and inpatient stays with length of stays (LOS) greater than 29 days, as that is the top 1% outlier.

## Medicare FFS Validation Data Set

We create the Medicare FFS validation data set from the 2014 MEDPAR file, which contains claims data for Medicare-covered inpatient stays at inpatient hospitals and skilled nursing facilities (SNFs). We first select all inpatient stays provided by facilities located in Colorado (the first two characters of the provider number equals "06"). Then we exclude "information-only" claims for enrollees in private Medicare Advantage plans. We further exclude claims for SNF stays and claims from long-stay hospitals, and exclude claims processed as secondary; with allowed amounts equal to zero; claims with an allowed amount greater than charges; and inpatient stays with outlier values for length of stay or allowed amounts.

## II. Algorithm

### Unit of analysis

As discussed in the main text, the unit of analysis we use for our algorithm is a *hospital-health plan-year-MSDRG* (a “quad”). This group is likely more fine-grained than actual contracts are, in multiple ways.

First, in reality, contracts are large, complex documents that cover the entirety of services and procedures hospitals provide. In this sense, semantically speaking, the unit of analysis we use may better be termed “subcontracts” in that they are part of the larger contract between a payer and a hospital. However, for ease of exposition we refer to them as contracts.

Second, we are likely being too fine-grained because anecdotal evidence suggests that hospitals and health plans neither (i) bargain over each individual (of the approximately 750) MS-DRGs that exist, nor (ii) renegotiate every year. While, the procedural code we use to define inpatient contracts is the MS-DRG, real contracts likely set prices at a broader group of services (such as Major Diagnostic Codes), and may only specify prices for a few select MS-DRGs, or even finer codes, like ICD-9 codes. Using MS-DRGs rather than a broader code gives the algorithm a higher probability of defining the contract correctly because it is narrow enough not to mix different payment mechanisms together (i.e. if we classify orthopedics as fixed rate there is no guarantee that all MS-DRGs within orthopedics are classified as fixed rate if they are negotiated individually). We are being similarly conservative by allowing

contracts to vary each year. Anecdotally, experts suggest that hospitals and insurance companies re-negotiate contracts approximately every two years.

While we could broaden the unit of analysis to instead, for example, define plan-hospital-year “triads” and use our algorithm to identify contract types within each triad, we would then need to change our tests for contract types. For example, our current algorithm looks for fixed-rate contracts by testing for uniform allowed amounts within a quad, and looks for per-diem contracts by testing for uniform ratios of allowed amounts to LOSs within a quad. With triads, we would not want to test for uniform allowed amounts (or uniform ratios of allowed amounts to LOSs) because allowed amounts would undoubtedly vary among at least some DRGs within a triad, so contracts could be misclassified as discounted charges or per diems depending on the case. For fixed rates, we could instead test for uniform multiples of Medicare’s MS-DRG weights, which could result in more claims and spending being classifiable, but would change the rest from simply looking for fixed rates to looking for a specific type of (i.e., a multiple-of-Medicare) fixed rate. And again if hospitals and health plans contract using different multiples of Medicare within a triad, then there would be risk of misclassification. Moreover, if hospitals and health plans contract using fixed rates, but with an APR-DRG grouper, or some idiosyncratic set of (non-Medicare based) relative rates, a triad-based algorithm would similarly be led astray. Our general preference is to be more conservative in the claims we treat as classifiable, and more definitive in our findings regarding contract types for those classifiable claims.

On the other hand, it is possible that our unit of analysis is too broad, which would be the case if contracts were negotiated based on individual procedures instead of MS-DRGs. However, first, as we mention above, anecdotal evidence and conversations with industry participants suggest that very few segments of contracts are negotiated over specific procedures and, second, as we discuss further later, the quality of our classifications are highly dependent on the number of observations within each quad (i.e. how many claims fall within each quad). The narrower our classifications are, the fewer claims we observe in each unit, and thus the fewer useful quads and less useable data we have. We discuss this in more detail later in

this Appendix. Moreover, our conversations with industry participants indicate that MS-DRG codes are frequently used in contracts.

There is also the possibility that hospitals and health plans do not negotiate contracts based on January-December calendar years, but rather on fiscal years with different start-months. In this case, and if there was a change in contract type in the middle of the year, our algorithm will either identify one type over another, or render the unit unclassified. We do see evidence of fiscal year contracts that do not start in January—i.e., our algorithm clearly identifies contracts that switch from one type to another on a given date. Future work will incorporate this feature of contracting into our algorithm.

Finally, we do not know whether negotiations happen at the health plan or payer family level – e.g., whether a given payer’s (say Aetna’s) PPO negotiates different contractual terms than that payer’s HMO. Therefore, as we did with procedural codes, we err on the side of caution by going too fine-grained and allowing contracts to vary at the health plan level. We discuss results from sensitivity checks we did around this element of our algorithm below, and in the main text.

#### Examples to illustrate how the algorithm works

Consider the example of a quad  $q$  with  $N_q = 100$  claims in it. The existence of such a quad would imply that our data contains 100 observations of patients having the same procedure (e.g., say, MS-DRG 775, an uncomplicated vaginal delivery), occurring in the same hospital (e.g., hospital A), in the same year (e.g., 2010), paid by the same health plan (e.g., payer 1). Now suppose the true contract type for that procedure was a pure discounted charge arrangement, meaning that hospital A and payer 1 contracted in 2010 to pay all incidences of MS-DRG 775 a certain percentage (e.g., 60 percent) of billed charges. Then, the discounted-charge construct,  $X_i$ , which equals the ratio of the allowed amounts to billed amounts should be exactly the same number (.6) for each patient  $i = 1, \dots, 100$  in the quad (each patient having an

uncomplicated vaginal delivery in 2010 at hospital A, paid by payer 1). Thus, we would expect the CV for construct (1) to be zero (or close to zero<sup>1</sup>) for the quad. In contrast, we would *not* expect all these patients' length of stays to be constant, which would introduce variance into construct (2), nor would we expect the allowed amount by itself to be constant across those patients either, because allowed amounts would likely vary based on the intensity of services provided to each patient within the MS-DRG, thus introducing variance into construct (3). Therefore, we expect our algorithm to find the discounted-charge construct as having the smallest COV among the three for this quad, and would thus label quad  $q$  and all claims therein as having a discounted-charge contract.

We provide specific examples of quads that our algorithm has identified as either: discounted charge, per diem, or fixed-rate contracts. We offer graphical depictions that illustrate how our formulae and algorithms work in each of the three contract type constructs.

*Examples of Discounted-Charge Contracts.* Appendix Figure 2 shows three examples (corresponding to panels a through c) of quads that our algorithm classified as discounted charges. Example 1 is a quad with 147 inpatient claims for MS-DRG 766 (Cesarean Section w/o cc/mcc) in year 2011. Example 2 is a quad with 365 inpatient claims for MS-DRG 775 (Vaginal Delivery w/o complicating diagnosis) in year 2011. Example 3 is a quad with 174 inpatient claims for MS-DRG 470 (Major Joint replacement) in year 2012. Each dot in each figure corresponds to one observation within the quad (many dots are overlaid which gives the impression of fewer observations).

Each example consists of two panels. The left panel plots prices against length of stays (LOS), and the right panel plots prices against charges. As these pictures demonstrate, in each of these examples quads, the right panels have less spread than the left panels, indicating that the relationships between prices and

---

<sup>1</sup> The reader should note that, because of contract complexities mentioned earlier, it is unlikely that any particular variance is exactly zero. However, using the decision rule to choose the contract type with the minimize variance helps to alleviate this concern.

charges are tight (see correlation coefficients in Figures) and significantly tighter than the relationship between prices and LOSs. Furthermore, there is substantial variation in allowed amounts. In the language of our algorithm, the CV of construct (1) is smaller than the CV of constructs (2) and (3), and thus we label these quads as being discounted-charges.

*Examples of Per-Diem Contracts.* Appendix Figure 3 provides three examples (corresponding to panels a through c) of quads that our algorithm labeled per diem. Examples 1 and 2 are for MS-DRG 766 (Cesarean Section w/o cc/mcc), while example 3 is for MS-DRG 795 (Normal Newborn); these three quads have 257, 284 and 152 claims, respectively.

Each of the three panels displays two graphs: the left panel plots prices against LOSs, and the right panel plots prices against charges<sup>2</sup>. Each dot in each of the scatter plots represents a single observation in the quad (many dots are overlaid which gives the impression of fewer observations). In these plots, prices plotted against LOSs (left panels) show strong, tight, positive linear relationships (see correlation coefficients reported in Figures) which indicates that claims for patients staying two days would be priced roughly two times that of patients staying one day. Plotting prices against charges reveals (right panels) these variables have a looser positive linear relationship (see correlation coefficients in Figures). Moreover, the presence of many allowed amounts in both panels suggests that there is no single fixed payment.

Our algorithm classifies these quads as per diem because the CVs corresponding to the discounted charge, per diem and fixed-rate constructs are minimized by the per-diem contract. In other words, the variance in the ratio of the allowed amounts to LOSs (construct (2)) is smaller than the variance of either the ratio of allowed amounts to billed charges (construct (1)), or the variance in allowed amounts (construct (3)).

---

<sup>2</sup> In all graphs, prices are masked to comply with our DUA.

*Examples of Fixed-Rate Contracts.* Appendix Figure 4 shows three examples (corresponding to panels a through c) of quads that our algorithm identifies as having fixed-rate contract types. As above, each of the three examples displays two panels: in the left panels we plot prices against LOSs, and in the right panels we plot prices against billed charges. Each dot in each scatter plot represents a single claim in the quad (see correlation coefficients in Figures); these three quads have 585, 114 and 228 claims, respectively.

In contrast with the per-diem contract figures we discussed before, prices plotted against either LOS or billed charges are roughly fixed at a single rate, as can be seen by the single, strong horizontal line that appears in each panel. This indicates that allowed amounts are constant regardless of variation in billed charges or LOS, as expected for fixed rate-contracts.

We label these quads as fixed rate, because the CVs corresponding to the three empirical constructs are minimized by the fixed rate CV. That is, there is little variance in allowed amounts (construct (3)), as compared with the ratios of either allowed amounts to billed charges (construct (1)) or allowed amounts to LOSs (construct (2)).

### Distribution of CVs across contract types

Appendix Figure 5 shows the distribution of our quads' three coefficients of variation (CVs). We display box-and-whisker plots of CVs for all three empirical constructs, depending on whether the quad was ultimately classified as discounted charge, fixed rate, or per diem. For example, Appendix Figure 5-a shows the distribution of the 3 CVs for quads classified as having discounted-charge contracts (N=3,744 quads, see Appendix Figure 6). For those quads, the CVs of the discounted-charge construct (allowed amount/charge amount) ranges from 0 to 1.61, with a median of 0.05, while the CVs of the fixed-rate construct (allowed amount) range from 0.003 to 3.76, with a median of 0.4, and the CVs of the per-diem

construct (allowed amount/LOS) range from 0 to 2.98, with a mean of 0.36. (Numbers from unpublished table.) Appendix Figure 5-b & c show analogous distributions of CVs across quads ultimately classified as fixed rates (N=1,976 quads, see Appendix Figure 6), and per diems (N=2,406 quads, see Appendix Figure 6).

Like Appendix Figures 2-4, Appendix Figure 5 also offers some intuition for how our algorithm works: the “winning” (i.e., lowest) CV among the 3 empirical constructs within a quad determines the classification of that quad. For example, the CV for the discounted-charge construct will necessarily be smaller for quads ultimately classified having discounted-charge constructs compared with CVs for the fixed rate or per-diem constructs.

#### Other contracting arrangements

While the three contract arrangements (discounted charges, per diems, and fixed rates) discussed in the main text are the most prevalent according to our reading of the literature and conversations with industry leaders, additional contract arrangements exist. For example, contracts may blend types (e.g., specifying fixed rates with an outlier provision based on discounted charges, or stop-loss provisions). As another example, implantable medical devices may be reimbursed through carve-out provisions, where they are reimbursed independently from the procedure performed. Certain ancillary services such as imaging and pharmaceuticals may also be reimbursed differently, although they may be more relevant to outpatient settings, which we do not study here. Finally, contracts might stipulate non-claims based quality bonuses or “infrastructure” payments that are not captured in the CoAPCD.

The fact that our algorithm simplifies contracts by focusing on the three main arrangements means that we will misclassify other contracting nuances as one of our three types unless we explicitly allow for a

fourth category to catch contracts that seemingly do not fit well into the discounted charge, per diem, or fixed rate frameworks. To alleviate this problem, we do a sensitivity analysis – the methods and results of which are discussed in more detail below – where we run the algorithm *with bounds*, which more conservatively places contracts with “too much” variance into an unclassified category. In addition to this sensitivity check, we also conduct thought experiments to understand where and how we might misclassify other contracting arrangements, and thus how this may affect our results. We discuss these below, when we discuss the results from our Medicaid validation sample. However, while our algorithm is not designed to detect other contract types, our simplified approach still allows us to assign claims to classifications that reflect similar characteristics of risk sharing and incentives.

### III. [Additional descriptive statistics to support results section](#)

In this section we provide additional characteristics of members, procedures, and payers in our analytic and validation data sets. Table 1 in the main text shows descriptive results in terms of total spending for each year (rows 1-6) and overall (last row), for the analysis sample (column 1) and for the Medicaid validation sample (column 7); Appendix Table 1 shows analogous results in terms of the number of claims. For the private sample, the general increase in claims and spending over time reflects increased coverage by the CO APCD. There was also an increase in claims and spending in the Medicaid sample in 2014, as Colorado’s Medicaid expansion came online on Jan 1 of that year.

Appendix Table 2 shows the distribution of patient ages and genders for both our private sample and Medicaid validation sample. Appendix Table 3 extends Table 2 to show the top-25 clinical conditions for the private and validation samples. We created most of the clinical condition groups by ranking MS-DRGs for the private sample, then taking top-ranked MS-DRG pairs that are the same except for their complicating conditions status. For example, the condition, “c-section,” includes the pair of MS-DRGs 765 (cesarean section with complicating condition or major complicating condition) and 766 (cesarean

section without complicating condition or major complicating condition). However, a few MS-DRGs do not have a paired match, such as normal newborn, while others groups include more than just the pair. For example, the group “vaginal delivery” includes MS-DRGs 774 (vaginal delivery with complicating conditions), 775 (vaginal delivery with complicating conditions), as well as 767 (vaginal delivery with sterilization and/or D&C) and 768 (vaginal delivery with OR procedure except sterilization and/or D&C). Vaginal deliveries are the most common clinical condition, accounting for 18.7% (26,200/140,000) of private stays and 8.5% (\$164m/\$1935.7m) of private spending. Lower extremity joint replacements account for 5.4% (7,500/140,000) of private claims, and 11.0% (\$213.2m/\$1935.7m) of private spending.

Regarding private payers, our final private sample contains 30 distinct payers, with an average of 4,668 claims observed per payer over the six-year time period. Appendix Table 4 shows the interquartile range of claims we observe per payer, per year. We cannot obtain additional details about payer characteristics because the data contain only de-identified payer codes; however, Appendix Table 4 lists all payers in our sample and their respective sizes in terms of number of claims observed. The Medicaid FFS sample has only one payer.

### Results by number of claims and quads

Recall that, because our algorithm relies on calculating the standard deviation of the empirical constructs (1), (2), and (3) within each quad, we must eliminate quads for which we cannot reliably measure variance – that is, those quads that have too few observations. In practice, this means we eliminate quads with fewer than three claims; we call those dropped claims *unclassifiable*. While our full inpatient data sample has 48,121 quads, the vast majority of these has only one or two observations, and thus ultimately must be dropped. After this exclusion, we are left with 8,229 quads that have at least three claims; however, as shown in the paper, these quads make up the majority of claims (66.5%) and spending

(55.5%) in the total sample. Appendix Table V provides more information about the size distribution of quads in terms of their number of claims and spending by year.

Appendix Tables 1 and 6 reproduce our main results that were presented in Table 1, but in terms of claims and quads instead of spending. The main patterns persist: In terms of the number of claims, the algorithm classifies 42.5% of classifiable claims as discounted charges, 31.9% of classifiable claims as fixed rates, and 24.9% of classifiable claims as per diems, which is fewer fixed rates and more per diems as compared to when the results are denominated in dollars. This indicates that lower dollar-value claims are paid per diems, and higher ones are paid fixed rate. The Medicaid and Medicare results also show more fixed rates – but also more discounted rates – when denominated in terms of number of claims, rather than spending.

In terms of quads, for the private sample, our algorithm classifies 3,703 (45% of classifiable) quads as having a discounted-charge contract, 1,959 (23.8% of classifiable) quads as having a fixed-rate contract, and 2,386 (29% of classifiable) quads as having a per-diem contract. One conclusion from this is that per-diem contracts are generally found among smaller quads, with smaller value claims, hence the reason that the fraction of per-diem contracts is smaller in terms of claims and spending. For the Medicare sample, our algorithm classifies 345 (4.9% of classifiable) quads as having a discounted-charge contract, 6,089 (86.4% of classifiable) quads as having a fixed-rate contract, and 536 (7.6% of classifiable) quads as having a per-diem contract. For Medicaid, our algorithm classifies 1,637 (11.6% of classifiable) quads as having a discounted-charge contract, 11,246 (79.7% of classifiable) quads as having a fixed-rate contract, and 1,101 (7.8% of classifiable) quads as having a per-diem contract.

### Results by additional clinical conditions

Appendix Table 3 extends Table 2 from the main analysis to show how contracts are distributed across MS-DRGs within clinical conditions, because the algorithm runs at the MS-DRG level (i.e., quads are

defined at the MS-DRG level). As can be seen in Appendix Table 3, there is some variation among MS-DRGs within a clinical condition. Among almost all clinical conditions, for instance, procedures “with complicating conditions” are more likely paid by discounted charges than their sister procedures “without complicating conditions.”

### Results by payer family versus health plan

Appendix Table 4-b shows variations in the distribution of contract types for each health plan within each (unidentified) payer family. As described in the main text, in our data set a health plan (which is a 4-digit payer code) is typically a different product line (e.g., HMO, PPO), but in a few cases, could be an employer or market segment (e.g., individual market). These payer codes are the codes under which plans submit their claims to the CO APCD. All health plans belonging to same payer family (e.g., Aetna, Cigna) have the same first 3 digits. We group our quads by the finer category of health plan, rather than payer family, to be conservative in how contracts are bargained and negotiated. Variation in the distribution of contract types within the same payer family exists but is small.

### Results Exploring the Role of Special Provision Payments in Medicare FFS

Appendix Table 11 shows a sequence of additional analyses we do to understand why, contrary to our expectations, our algorithm does not classify 13.7% of Medicare spending as fixed rates. The first row of Appendix Table 11 (Main Analysis) corresponds to our main Medicare FFS result in the paper, showing we classify 86.3% of spending as fixed rates. The second row, Additional Analysis 1, uses the same Medicare FFS sample as the main analysis but uses DRG\_PRICE as the payment variable instead of the Allowed Amount. As per the ResDAC website, the DRG\_PRICE is the amount that would have been paid if no deductibles, coinsurance, primary payers or outliers were involved. It is the outlier payments

that are important here. As shown in Appendix Table 11, removing outlier payments reduces “misclassification” of discounted charges by 6.9 percentage points (between the Main Analysis and Additional Analysis 1, classification of fixed rates increases from 86.3% to 93.2% of total spending, while discounted charges drops from 7.8% to 1.4%). Additional Analysis 2 also uses DRG\_PRICE as the payment variable, but now also omits CAHs, thus showing how much of the “misclassification” of per diems is because of CAHs (between Additional Analysis 1 and Additional Analysis 2, classification of fixed rates increases from 93.2% to 95.2% of total spending, while per diems drops from 5.2% to 3.2%). Additional Analysis 3 is the same as Additional Analysis 2, but now also omits special units, thus showing how much of the “misclassification” is because of special unit claims, such as rehabilitation and psychiatric claims (from Additional Analysis 2 to Additional Analysis 3, classification of fixed rates increases from 95.2% to 98.6% of total spending). Additional Analysis 4 is the inverse of Additional Analysis 2, keeping only CAHs and using DRG\_PRICE, thus showing that CAHs are primarily (98.3%) paid by per diems.

#### IV. Validation and Sensitivity Analyses

We perform three sets of analyses to check the robustness of our algorithm and our findings. First, as a sensitivity test, we exclude quads with fewer than four claims, instead of fewer than three (see Appendix Table 7). We do this because our measure of variance (the CV) within each quad may be noisy in quads with so few observations. Second, we check whether our trends reflect compositional changes in payers in the CO APCD over time (Appendix Table 8). Finally, to address the issue that our algorithm may be ignoring contracting complexities by forcing all quads into one of three categories, we do a validation check we call the *algorithm with bounds*. There, we add two bounds to the basic algorithm, such that we only classify quads when the (i) difference between the CV of the “winning” construct and the CV of the “runner-up” construct exceeds a minimum bound, and (ii) “winning” construct’s CV is not too far from zero (Appendix Tables 9 and 10). This effectively treats noisy quads as not classified.

Appendix Table 7 shows results for both our private and Medicaid samples when we exclude quads with fewer than 4 claims, instead of fewer than 3. We present results in terms of claims (7-a), spending (7-b), and quads (7-c). Our findings do not change significantly, except slightly more spending is classified as fixed-rate contracts among the private sample. This may support a hypothesis that our algorithm performs better when there are more claims per quad, if fewer claims leads to more noise and more likelihood of classifying as discounted charge. However, it may also indicate that sparser quads are simply more likely to be discounted charges.

Appendix Table 8 shows results from sensitivity analyses where we check to ensure that compositional changes in payers that the CO APCD added over time are not confounding our results about the changing contract classification mix over time. In particular, we re-run our analysis on the private sample, excluding one very large payer, who entered the database in 2011; their claims account for roughly one-third of our total sample. Our results about trends, especially concerning the declining prevalence of

discounted-charge contracts, remain even without this payer (which seems to have predominantly fixed-rate contracts).

Finally, we impose bounds on our algorithm to check that we are not ignoring contracting complexities by effectively forcing all quads into one of three categories. We redo the analysis with the restrictions that the chosen contract has: (i) a CV below 0.05, and (ii) a CV that is 0.05 below the next smallest CV. The results (Appendix Table 9) show that more contracts are deemed unclassifiable, as expected, but the proportion of different contract types remains substantively unchanged. In Appendix Table 10, we show how the classification results (including the number of unclassifiable) from the algorithm with bounds changes as we use different values for the bounds, ranging from 0.01 to 0.1. As can be seen, the minimum bound (i.e., the one requiring the CV to be below a certain threshold) is more likely to bind, relatively to the bound requiring minimum distance between the “winning” CV and the “runner-up.”

We also implement an additional algorithm in which we bootstrap sample each quad for 100 repetitions. We then classify each quad if our initial algorithm produces the same response in at least 90% of the repetitions. Our results (author tables) are similar to other descriptions of our results. However, one weakness of this approach is that it effectively eliminates all quads with only 3 claims. In expectation, the bootstrap sample will produce quads with identical claims a significant proportion of the time. Quads with too few claims do not produce variation that allow us to differentiate contract types, which drastically reduces the probability that the algorithm will find 90% agreement across 100 repetitions.

#### V. [More evidence that discounted-charge contracts exist](#)

The upper left pane of Appendix Figure 7 shows the distribution of claims that our algorithm identified as discounted charge over the range of reimbursement rates (i.e., the ratio of the allowed amount to the charge amount, or, empirical construct (1)). There are discrete spikes at round numbers (those divisible by .05), which we take as evidence of discounted-charge contracts being negotiated around cognitive

benchmarks. In contrast, the distribution of claims that our algorithm identified as having per-diem or fixed-rate contracts is smooth over the range of reimbursement rates, indicating that our algorithm is likely classifying these claims correctly. Looking at the lower right panel of Appendix Figure 7, we see a mix of these two findings for our not classified claims, which suggest we are not classifying some of each type of contract, or that there exist more complicated, hybrid contract structures for which our simple algorithm cannot account for.