

American Journal of Advanced Technology and Engineering Solutions

Volume: *12;* ***Issue:*** *09*
Pages: *01-40*
eISSN: 3067-0470

open access
Crossref
PEER-REVIEWED

Federated Learning and Differential Privacy for Cross-Firm Construction-Materials Supply-Chain Disruption Forecasting: A Quantitative Analysis of Privacy-Utility Trade-Offs

Asheke Mostofa[1]; Tashriq Ahmed Bin Azad[2];

[1]. *Supply Chain, Engineering Management & Manufacturing Researcher; M.S. in Textile Engineering, Wuhan Textile University, China*
Email: asheke.mbstu@gmail.com

[2]. *B.Sc. in Civil and Environmental Engineering, North South University, Dhaka, Bangladesh;*
Email: tashriqahmedbin.azad@snhu.edu

Doi: 10.63125/g4at4237
Policy Page DOI: 10.63125/ajates-crossmark-policy

Received: *12 June 2022;* ***Revised:*** *09 July 2022;* ***Accepted:*** *13 August 2022;* ***Published:*** *02 September 2022;*

Abstract

Construction firms forecast materials disruptions from their own order histories, because delivery records reveal suppliers, volumes and vulnerabilities to competitors, yet no single firm's short history can estimate the regional component of disruption risk. Federated learning with differential privacy promises a shared forecaster without shared records, but its costs for consortia with short, correlated order panels, rare outcomes and a common regional shock are unmeasured. This study quantifies the trade-offs in a controlled simulation with a fully specified generator of order-line panels: federations of 5 to 100 firms, 6 to 24 periods of history, a shared regional factor and three levels of heterogeneity trained logistic regression and neural network forecasters of next-period disruption under local-only, centralised, federated, sample-level differentially private (ε 0.5 to 8) and client-level differentially private (ε 4 to 64) regimes across 1,840 runs, with Rényi-DP accounting, out-of-time evaluation (AUC, calibration, watch-list capture, expected cost) in a base and a shock scenario, and a membership-inference attack. Federation recovered nearly all of the pooling benefit (0.3 AUC points below centralised training at moderate heterogeneity) and most helped small firms and short histories; under strong heterogeneity, local fine-tuning of the shared forecaster beat both shared and local models. Sample-level privacy cost the logistic forecaster 1.7 points at ε = 4 and 4.2 at ε = 1 with ten periods of history, less with longer ones, and crippled the network; client-level privacy at conventional budgets was unattainable with ten firms (19.6 points at ε = 4). A regional shock at the horizon left ranking almost intact but cut decision value through calibration (FedAvg saving 33% to 24%; local 26% to 13%); the federated forecaster kept more of its value than local ones, and local recalibration did not help. The attack gained nothing measurable against the shared logistic forecasters but did against isolated local models. The findings are translated into design guidance, a governance workflow, a decision tree and a roadmap. Limitations are synthetic data, small models and a single black-box attack.

Keywords

Construction Supply Chain; Materials Disruption Forecasting; Federated Learning; Differential Privacy; Privacy-Utility Trade-Off; Regional Supply Shock; Membership Inference; Simulation Study

INTRODUCTION

Construction projects depend on the timely arrival of materials from supply chains that are fragmented, project-based and loosely coupled: a contractor buys cement, aggregates, structural steel, timber, gypsum board, insulation and glazing from many suppliers and distributors through short relationships that re-form with every project (Vrijhoef & Koskela, 2000; Dubois & Gadde, 2002; Behera et al., 2015). Late, short or absent deliveries are among the most frequently cited causes of delay and cost overrun in construction wherever the question has been asked, from large projects in the Gulf and South-East Asia to public housing programmes (Assaf & Al-Hejji, 2006; Sambasivan & Soon, 2007; Le-Hoai et al., 2008; Hwang et al., 2013), and materials price volatility and inflation feed directly into project budgets (Musarat et al., 2021; Ng et al., 2004). The pandemic years showed how quickly a regional supply shock propagates across the sector: surveys of contractors in the United States reported materials shortages, price surges and delivery delays as early impacts, affecting firms that had never dealt with the suppliers in question (Alsharef et al., 2021; Assaad & El-adaway, 2021). Forecasting which material lines will be disrupted in the coming weeks, so that buffers, substitutions or alternative sources can be arranged, is therefore a decision that every contractor and builders' merchant faces, and supervised learning on order histories is the natural tool (Bilal et al., 2016; Gondia et al., 2020; Pan & Zhang, 2021).

The obstacle is data. A single firm sees only its own order lines, disruptions are infrequent, and histories are short because projects are short; the information that would make a forecast better, the delivery performance of the same suppliers and the same material categories as experienced by other firms in the same region, sits behind the confidentiality boundaries of competitors (Titus & Bröchner, 2005; Cheng et al., 2010). The construction supply-chain literature has documented for two decades that information flows between firms are limited, that collaboration is hindered by adversarial relationships, short-term contracting and the fear of opportunism, and that the data a firm holds about its suppliers reveal its sourcing, prices and vulnerabilities (Akintoye et al., 2000; Xue et al., 2007; Aloini et al., 2012; Meng, 2012). The general operations literature gives the same account: shared information is valuable but is withheld when it can reach competitors or be used strategically by the common party (Lee et al., 2000; Li, 2002; Ha & Tong, 2008; Fawcett et al., 2008). Regional shocks make the loss acute, because the regional component of disruption risk is exactly what no single firm's history can estimate from its own few records.

Federated learning is the arrangement under which firms could learn from one another's experience without exchanging records: each firm trains a shared forecasting model on its own order lines and sends only model updates to a coordinator that averages them (McMahan et al., 2017; Kairouz et al., 2021; Yang et al., 2019a). Model updates and the released model nevertheless leak information about the training records through membership and property inference and through gradient inversion (Shokri et al., 2017; Melis et al., 2019; Nasr et al., 2019; Zhu et al., 2019), and differential privacy is the mechanism that bounds what any such attack can learn, at a price in accuracy that depends on the privacy budget, on the number of participants, on the data each holds and on whether the guarantee covers single records or whole participants (Dwork & Roth, 2014; Abadi et al., 2016; McMahan et al., 2018; Wei et al., 2020). For a group of construction firms these trade-offs are the design problem, and they have not been measured for this setting: the federated learning literature evaluates on image and text benchmarks with thousands of clients (McMahan et al., 2017; Li et al., 2020b), the construction analytics literature works with single-firm or single-project data and no privacy mechanism (Gondia et al., 2020; Yaseen et al., 2020; Darko et al., 2020), and a companion study of supplier-risk assessment across enterprises measured the trade-offs for a cross-sectional risk score rather than for a forecast under regional shocks (Aditya & Ahmed, 2022; Kazi Mohammad Khalid & Sumaiya Sadmeen, 2022; Mansura Akter & Sumaiya, 2022). A cross-firm materials federation has features that change the answer: few participants (tens of firms), short panels of order lines with correlated records, rare outcomes, heterogeneity in material mix and supplier base, a shared regional factor that produces simultaneous disruptions, and a decision use, arranging buffers or alternatives for the lines flagged, whose costs are asymmetric (Mashiat & Md. Morshedul, 2022; Md Imran & Jannatul, 2022; Md Maksudul & Md Tazul Islam, 2022).

This study quantifies the privacy-utility trade-offs for that setting with a controlled simulation. It

constructs federations of firms holding synthetic order-line panels whose generating process is fully specified, including a regional market factor shared by all firms, trains next-period disruption forecasters under local-only, centralised, federated and differentially private federated regimes, and measures utility (discrimination, calibration, watch-list capture and expected cost at three cost ratios), its distribution across firms, its robustness when a regional shock arrives at the forecast horizon, and empirical privacy against a membership-inference attack. It addresses four research questions:
- RQ1: How much forecasting utility does federated training recover relative to local-only training and lose relative to centralised pooling, and how does this depend on heterogeneity across firms, on the number of firms and on the length of their order histories? - RQ2: How does utility change with the privacy budget under sample-level and under client-level differential privacy? - RQ3: How is the utility loss distributed across firms of different sizes, and how do the regimes compare when a regional supply shock arrives at the forecast horizon, including the decision value at different cost ratios? - RQ4: How much does a membership-inference attack gain against the released models with and without differential privacy, and how does this empirical advantage compare with the formal guarantee?
The contribution is threefold. First, the study gives quantitative trade-off curves for a setting the literature has not measured: small federations of construction firms with short, correlated, imbalanced order panels and a shared regional shock process. Second, it shows that sample-level differential privacy is affordable for a low-dimensional forecaster whereas client-level guarantees at conventional budgets are not attainable in federations of realistic size, that privacy noise and regional shocks both degrade decision value through calibration faster than they degrade ranking, and that the federated forecaster retains more of its decision value under a shock than isolated local models because it has learned the regional component from every member's history. Third, it translates the results into design guidance, a governance workflow, a decision tree for choosing the mechanism and an implementation roadmap for a materials-supply consortium. Section 2 reviews the background. Section 3 describes the simulation. Section 4 reports the results. Section 5 presents the design guidance. Section 6 discusses the findings and limitations, and Section 7 concludes.

BACKGROUND

Construction-materials supply chains and their disruptions

Construction supply chains differ from manufacturing supply chains in ways that matter for forecasting. Production happens on a temporary site, the supply chain converges on that site and is reconfigured for every project, and the parties are tied by short contracts rather than long relationships; the result is a loosely coupled system in which coordination across firm boundaries is weak and learning between projects is lost (Vrijhoef & Koskela, 2000; Dubois & Gadde, 2002; Segerstedt & Olofsson, 2010). Reviews of construction supply-chain management describe four roles for it, from improving the interface between site and supply chain to integrating the chain, and list the implementation risks that fragmentation, adversarial culture and poor information flow create (Vrijhoef & Koskela, 2000; Behera et al., 2015; Aloini et al., 2012). Performance measurement frameworks adapted from the SCOR model show that construction logistics suffer from delivery reliability and lead-time problems that are visible only when measured across the chain (Thunberg & Persson, 2014), and materials management studies document the cost of poor material availability on site, the trade-off between storage and just-in-time delivery, and the practices that reduce shortages (Polat & Arditi, 2005; Said & El-Rayes, 2011; Caldas et al., 2015).
The consequences of materials disruption are well documented in the delay literature. Shortages, late deliveries and price changes of materials appear among the leading causes of delay in large projects in Saudi Arabia, Malaysia, Vietnam and elsewhere, alongside contractor financing, design changes and labour (Assaf & Al-Hejji, 2006; Sambasivan & Soon, 2007; Le-Hoai et al., 2008), and schedule performance of public housing projects depends on material and supplier factors as much as on management factors (Hwang et al., 2013). Materials costs move with construction cost and tender price indices whose forecasting has its own literature, from regression and time-series models of tender prices to multivariate models of the Engineering News-Record indices (Ng et al., 2004; Ashuri & Lu, 2010; Hwang, 2011; Shahandashti & Ashuri, 2013), and inflation in materials prices is a recognised source of budget overrun (Musarat et al., 2021). Resilience research in construction has turned from these consequences to the vulnerabilities and capabilities of the supply chain: extreme weather exposes

small and medium construction firms and their suppliers (Wedawatta et al., 2010), vulnerabilities and capabilities interact dynamically in the Malaysian supply chain (Zainal Abidin & Ingirige, 2018), and industrialised construction in Hong Kong faces supply-chain vulnerabilities whose management requires specific capabilities (Ekanayake et al., 2021b; Ekanayake et al., 2021a). The pandemic made the regional character of materials shocks visible: early United States surveys reported shortages, price increases and delivery delays across firms and trades at once (Alsharef et al., 2021), and guidance for the sector emphasised supply-chain monitoring and alternative sourcing (Assaad & El-adaway, 2021). Three properties of this setting shape the forecasting problem. First, disruption risk has a line-specific component (the state of a particular supplier and material) and a regional component (market tightness, transport, weather, demand surges) that is common to every firm in a region; the regional component is what a single firm with a short history cannot estimate, and it is also what produces the correlated failures that hurt most. Second, the records are panels: each line is observed repeatedly, so that records are correlated within a line and the effective sample size is smaller than the record count. Third, the decision is asymmetric: a missed disruption stops work or forces expediting and substitution, whereas a false alarm costs a buffer or an enquiry, so that the model is judged by expected cost at a threshold rather than by accuracy.

Analytics and machine learning in construction supply chains

Data analytics has entered construction management through big data on projects, sites and supply chains (Bilal et al., 2016), and reviews of artificial intelligence in construction engineering and management record a rapid growth of machine-learning applications, concentrated on cost, schedule and safety prediction (Pan & Zhang, 2021; Darko et al., 2020). Delay risk has been predicted with decision trees and naive Bayes classifiers on project records (Gondia et al., 2020) and with hybrid models (Yaseen et al., 2020); supply-chain relationship traits have been linked to project performance by structural models (Kim & Nguyen, 2018); and web-service architectures for supply-chain integration in construction have been proposed to connect the parties' systems (Cheng et al., 2010). In the general supply-chain literature, machine learning predicts supplier disruptions from a single firm's data (Brintrup et al., 2020), and reviews of artificial intelligence in supply chain risk management describe the shift from expert scoring to supervised learning (Baryannis et al., 2019). In every one of these studies the model learns from one organisation's records; the experience of the other buyers of the same suppliers is not used, and the privacy of the records is not a design parameter (Md. Morshedul & Aminul Islam, 2022; Syed Feroz & Md Mahamud Pervaz, 2022; Syed Feroz & Shabbir, 2022).

The statistical problem is one of rare-event classification with asymmetric costs, and the evaluation conventions of credit scoring apply (Lessmann et al., 2015; Baesens et al., 2003): discrimination by the area under the receiver operating characteristic curve (Hanley & McNeil, 1982; Bradley, 1997), calibration by the Brier score (Brier, 1950), and decision value by expected cost at the cost-optimal threshold (Elkan, 2001; Hand, 2009). Because the forecast is made at a point in time for the next period, evaluation must be out of time, on records after the training window, and the shock case, a period in which the regional factor moves sharply, must be evaluated separately, since a model that ranks lines well in calm periods can be badly calibrated when the base rate jumps.

Information sharing and confidentiality between construction firms

The value and the limits of information sharing are long-standing themes of both the construction and the operations literatures. Early surveys of the United Kingdom construction industry found limited collaboration and information exchange between contractors and suppliers, with trust and commercial sensitivity as the obstacles (Akintoye et al., 2000); studies of information flow in construction supply chains show how fragmented and transactional that flow is (Titus & Bröchner, 2005); and coordination mechanisms for construction supply chains in the Internet environment, from information sharing to joint planning, remain conditional on relationship quality (Xue et al., 2007; Meng, 2012). The operations literature explains the withholding: when firms compete horizontally, information shared vertically can leak to competitors through the common party, and the equilibrium is often no sharing or limited sharing (Li, 2002; Shang et al., 2016); contracts and trust can restore some of it (Ha & Tong, 2008; Anand & Goyal, 2009; Dyer & Singh, 1998), and surveys of supply chain collaboration list confidentiality, fear of opportunism and lack of trust among the main barriers (Fawcett et al., 2008; Kembro et al., 2014; Klein & Rai, 2009). Order-line histories are among the most sensitive data a contractor holds: they reveal

which suppliers it uses, at what volumes and prices, how reliably they deliver and where the firm is vulnerable. Any method that learns across firms must protect two things: the individual records (a particular line's history with a particular supplier) and the firm-level aggregates (which suppliers, what volumes, what disruption rates). The distinction reappears below as the distinction between sample-level and client-level differential privacy.

Privacy-preserving collaboration and differential privacy

Before federated learning, three families of methods allowed computation across confidential data sets. Anonymisation releases transformed records but is vulnerable to re-identification and destroys much of the data-mining utility of detailed records (Sweeney, 2002; Narayanan & Shmatikov, 2008; Brickell & Shmatikov, 2008); reconstruction attacks show that too many accurate aggregates reveal the records themselves, which motivated differential privacy (Dinur & Nissim, 2003; Dwork et al., 2017). Secure multiparty computation and homomorphic encryption compute a specified function without revealing the inputs (Yao, 1986; Lindell & Pinkas, 2009; Gentry, 2009; Paillier, 1999), and have been applied to supply chain planning and benchmarking (Atallah et al., 2003; Clifton et al., 2008; Pibernik et al., 2011; Kerschbaum et al., 2011), but training a forecasting model is an iterative computation for which fully cryptographic solutions remain expensive and whose output can itself leak. Distributed ledgers record provenance rather than protect data (Saberi et al., 2019; Kshetri, 2018).

Federated averaging trains a model across clients that keep their data: the server broadcasts the model, each client runs local stochastic gradient descent, and the server averages the returned updates weighted by data size (McMahan et al., 2017). The cross-silo form, in which the clients are organisations with larger data sets and reliable connections, is the one relevant to firms (Kairouz et al., 2021; Yang et al., 2019a), and applications have been reviewed in industry, finance and health (Li et al., 2020a; Yang et al., 2019b; Hao et al., 2020; Lu et al., 2020; Rieke et al., 2020; Nguyen et al., 2021; Khan et al., 2021). Statistical heterogeneity across clients slows and biases the averaged model, and proximal terms, control variates, corrected aggregation and personalisation address it (Li et al., 2020d; Hsieh et al., 2020; Li et al., 2020c; Karimireddy et al., 2020; Wang et al., 2020; Smith et al., 2017; Fallah et al., 2020; Mohri et al., 2019). Secure aggregation hides individual updates from the server (Bonawitz et al., 2017; Bonawitz et al., 2019), robust aggregation defends against faulty or malicious clients (Blanchard et al., 2017; Bhagoji et al., 2019; Bagdasaryan et al., 2020), and surveys catalogue the security and privacy threats (Mothukuri et al., 2021). Secure aggregation does not bound what the aggregate or the final model reveals, which is the role of differential privacy.

Differential privacy bounds the influence of any single record on the output of a randomised computation by (ε, δ) (Dwork et al., 2006; Dwork & Roth, 2014; Dwork, 2008); compositions of private computations consume a budget whose growth is tracked by advanced composition and by Rényi differential privacy (Dwork et al., 2010; Kairouz et al., 2017; Mironov, 2017). Differentially private stochastic gradient descent clips each example's gradient and adds Gaussian noise, with the moments accountant and its Rényi successors tracking the budget under subsampling (Song et al., 2013; Abadi et al., 2016; Wang et al., 2019); objective perturbation and private empirical risk minimisation give the dimension dependence of the utility cost (Chaudhuri et al., 2011; Bassily et al., 2014). In federated learning, DP-SGD at each client gives sample-level guarantees, whereas clipping and noising whole client updates gives client-level guarantees that have been shown to work with thousands of clients and to depend strongly on their number (McMahan et al., 2018; Wei et al., 2020; Truex et al., 2019; Agarwal et al., 2018); hybrid, local and personalised variants trade the guarantees against utility in other ways (Truex et al., 2020; Zhao et al., 2021; Hu et al., 2020; Triastcyn & Faltings, 2019). The choice of ε is a policy decision with economic framings rather than a single value (Hsu et al., 2014; Lee & Clifton, 2011; Ghosh et al., 2012), and the utility cost of privacy falls disproportionately on under-represented groups and small contributors (Bagdasaryan et al., 2019).

Inference attacks and the measured trade-off

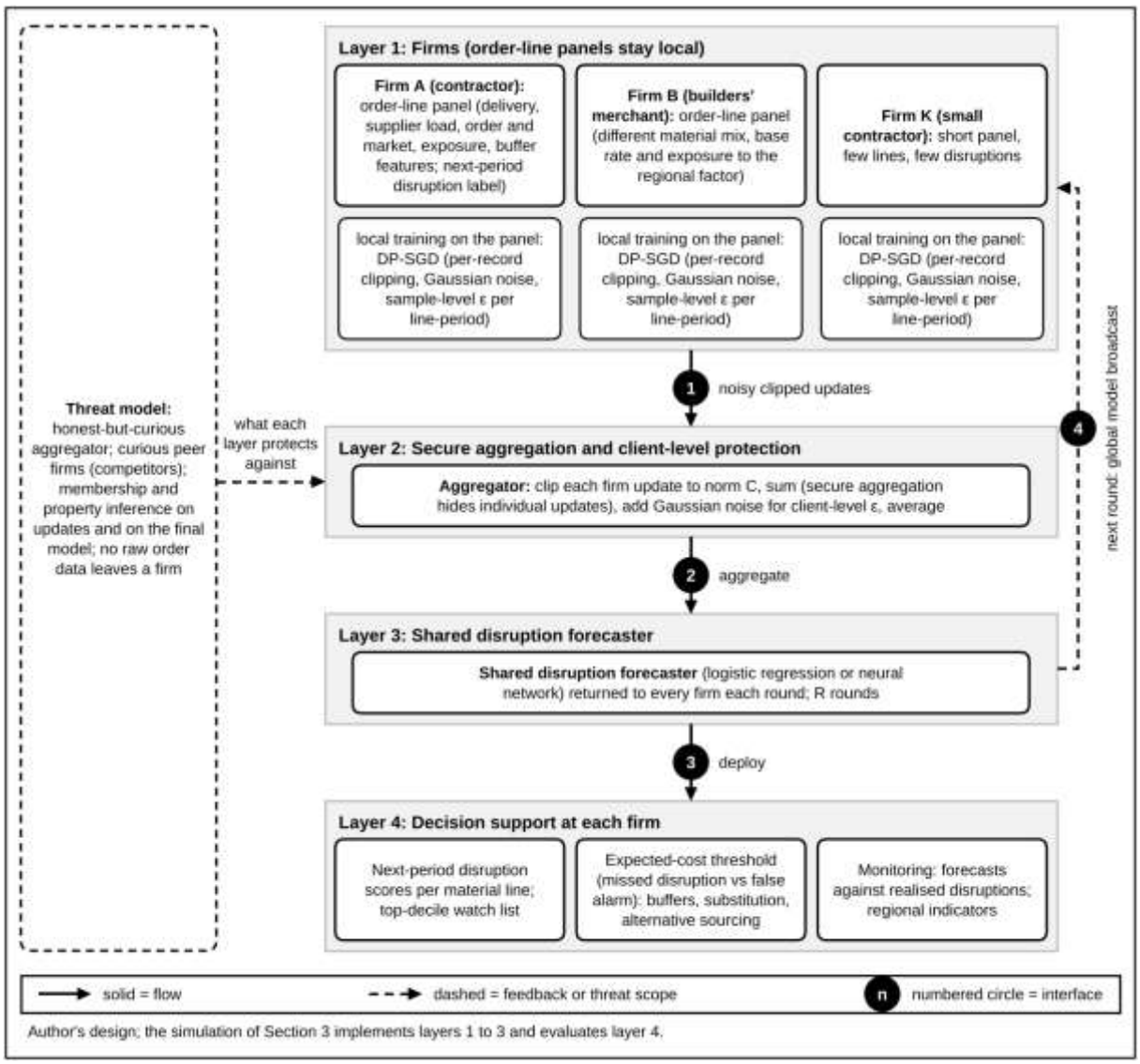
Membership inference decides whether a record was in the training set from the model's outputs and succeeds in proportion to overfitting (Shokri et al., 2017; Yeom et al., 2018; Salem et al., 2019; Sablayrolles et al., 2019; Song & Mittal, 2021); attribute inference and model inversion recover sensitive attributes (Fredrikson et al., 2015; Ateniese et al., 2015); and in federated learning, participants or the server can run white-box attacks on exchanged updates and reconstruct training examples from

gradients (Nasr et al., 2019; Melis et al., 2019; Zhu et al., 2019; Geiping et al., 2020). Empirical evaluations have compared the formal ε with the attacks' success: relaxed budgets leaked to membership inference in some settings while tight budgets cost much accuracy (Jayaraman & Evans, 2019; Rahman et al., 2018), and audits that estimate lower bounds on ε from attacks find the gap between measured and formal bounds to depend on the attack's strength (Jagielski et al., 2020). The trade-off is three-sided, formal budget, empirical advantage and utility, and it depends on the data and the model. Table 1 summarises the approaches to confidential cross-firm learning with what each protects and at what cost, and Figure 1 shows the architecture the simulation implements: firms that keep their order-line panels and train locally with sample-level DP-SGD, an aggregator that applies secure aggregation and optional client-level noise, a shared disruption forecaster, and decision support at each firm.

Table 1: Approaches to confidential cross-firm learning: what each protects, against whom, and at what cost

Approach	What is protected	Against whom	Cost or limitation	References
Contractual data sharing between firms	Nothing technically; relies on contracts and trust	Third parties	Leakage to competitors and suppliers; strategic withholding; weak collaboration culture in construction	(Akintoye et al., 2000; Li, 2002; Fawcett et al., 2008; Aloini et al., 2012)
Anonymised release of order histories	Direct and quasi-identifiers of released records	Recipients of the release	Re-identification of sparse records (few suppliers per material); large loss of utility	(Sweeney, 2002; Narayanan & Shmatikov, 2008; Brickell & Shmatikov, 2008)
Secure multiparty computation, homomorphic encryption	Inputs of a specified joint computation	Other parties and the operator	High cost for iterative training; output itself may leak	(Yao, 1986; Lindell & Pinkas, 2009; Gentry, 2009; Atallah et al., 2003; Clifton et al., 2008; Pibernik et al., 2011)
Federated learning	Raw order-line panels never leave the firm	Aggregator and peers see only updates	Updates and models leak through inference attacks	(McMahan et al., 2017; Kairouz et al., 2021; Melis et al., 2019; Nasr et al., 2019)
Federated learning with secure aggregation	Individual updates hidden from the aggregator	Aggregator	Aggregate and final model still unbounded; no formal guarantee	(Bonawitz et al., 2017; Truex et al., 2019)
Sample-level differential privacy (DP-SGD at each firm)	Each line-period record, with formal (ε, δ) bound	Aggregator and peers	Noise grows with fewer records per firm; calibration degrades; line-level guarantee needs group privacy	(Abadi et al., 2016; Wei et al., 2020; McMahan et al., 2018)
Client-level differential privacy (noised firm updates)	Each firm's whole panel, with formal bound	Aggregator (with secure aggregation) and peers	Noise relative to signal scales with 1/K; needs many participants	(McMahan et al., 2018; Wei et al., 2020; Agarwal et al., 2018)
Local differential privacy	Each record at the source, no trusted party	Everyone	Largest utility loss	(Kasiviswanathan et al., 2011; Duchi et al., 2013; Truex et al., 2020; Zhao et al., 2021)

Figure 1: Architecture of federated materials-disruption forecasting across construction firms, with the two privacy layers and the threat model



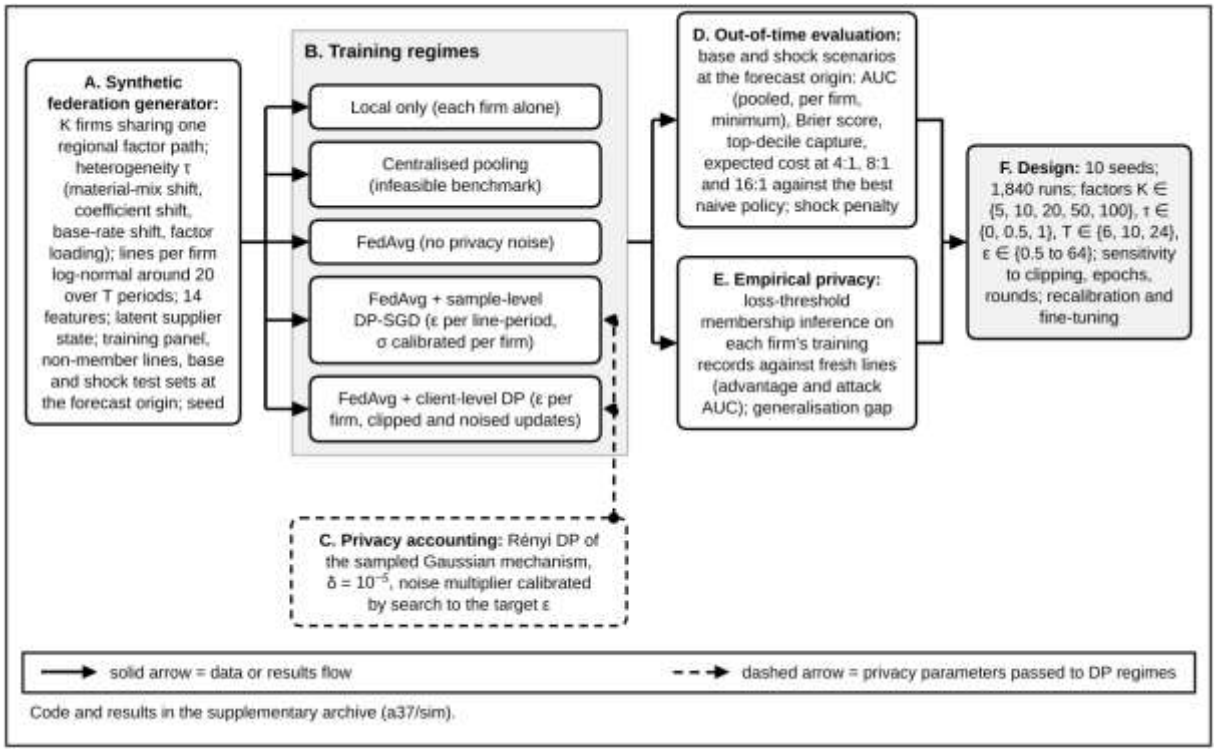
METHODOLOGY

Research design

The study is a controlled simulation experiment. Order-line histories of several construction firms with disruption labels cannot be assembled for research precisely because of the confidentiality constraints that motivate the work, and no public data set holds them. A simulation with a fully specified generating process allows the factors of interest (heterogeneity across firms, number of firms, length of history, privacy budget, privacy mechanism, arrival of a regional shock) to be varied with known ground truth, which observational data would not permit (Kleijnen, 2005; Sargent, 2013). The price is that the absolute numbers describe the simulated world; the design therefore sets the generator to realism targets taken from the construction delay and materials literature and from the companion supplier-risk study, reports the attainable (Bayes) performance as a reference, and interprets differences between regimes rather than absolute levels. The protocol (file protocol.md in the supplementary archive) fixed the generator, the models, the privacy mechanisms, the metrics and the design before the experiments were run. Two deviations are recorded in the protocol: the generator's intercept, regional effect, shock size, default history length and mean number of lines were adjusted once after a pilot that gave a prevalence of about 19 per cent, a shock prevalence of about 50 per cent and no gap between local and federated training, before any experiment of the grid was run; and the grid was started twice

by mistake, so that duplicate records of identical, seed-deterministic runs were removed before analysis. Figure 2 shows the pipeline.

Figure 2: Experimental pipeline of the simulation study: panel generator with a regional factor, training regimes, privacy accounting, out-of-time evaluation in base and shock scenarios, membership inference and design



Synthetic panel generator

A federation consists of K firms (contractors or builders' merchants) in one region. Each firm k holds a panel of m_k material lines, a line being one material category bought from one supplier, observed over T monthly periods; m_k is drawn from a log-normal distribution around a mean of 20 lines with a standard deviation of 0.6 on the logarithmic scale and a floor of 8, so that a federation combines a few large buyers with many small ones. A record is the feature vector of a line at the end of period t and the label is whether the line is disrupted in period t+1 (delivered late beyond the contractual window, short-shipped or not delivered). Two latent processes drive the panel. A line-specific supplier state s follows a first-order autoregression with persistence 0.7 and stands for the condition of that supplier and material (backlog, capacity, financial stress). A regional market factor f, shared by every firm in the federation, follows a first-order autoregression with persistence 0.6 and innovation standard deviation 0.8 (stationary variance one) and stands for regional materials tightness, price surges, transport and demand booms; each firm has its own loading on it, which represents its exposure through its material mix and import share.

Each record has fourteen standardised features in five correlated blocks (Table 2): delivery performance (lead-time deviation, lead-time variability, on-time rate), supplier load (supplier backlog, capacity utilisation, supplier financial stress), order and market (order-size ratio, three-month price change, regional demand index), exposure (import share, transport risk, single sourcing) and buffers and season (inventory cover, season hazard). Features load on the supplier state (delivery and load features) and on the regional factor (the delivery features more weakly, backlog, capacity, order size, financial stress, inventory cover, price change and above all the regional demand index), and carry block-correlated noise. The disruption logit for period t+1 contains a linear term in the period-t features with fixed population coefficients whose signs follow the construction supply-chain literature (late and variable deliveries, loaded and stressed suppliers, large orders, price surges, import exposure,

transport risk, single sourcing and regional demand raise risk; on-time history and inventory cover lower it (Assaf & Al-Hejji, 2006; Sambasivan & Soon, 2007; Caldas et al., 2015; Said & El-Rayes, 2011; Wedawatta et al., 2010)), two interactions (capacity utilisation with order size; import share with price change), a saturating term in the regional demand index, two threshold effects (on-time rate more than one standard deviation below the mean; inventory cover more than one standard deviation below), the firm's loading times next period's regional factor, and next period's supplier state. The forecaster observes only the period-t features; the innovations of the two latent processes between t and t+1 are unobservable, bound the attainable discrimination, and are the formal counterpart of the unforecastable part of a disruption.

Heterogeneity across firms is governed by one parameter τ: each firm draws a feature-mean shift (0.6 τ standard deviations per feature, its material mix), a coefficient shift (0.35 τ per coefficient), a base-rate shift (0.5 τ on the logit) and a loading on the regional factor of 1 + 0.4 τ times a standard normal draw (floored at 0.2), so that τ = 0 gives identically distributed firms with equal exposure, τ = 0.5 moderate differences (as between contractors of different trades in one region) and τ = 1 strong differences (as between a steel fabricator, a house builder and a civil contractor). For each firm the generator draws the training panel (periods 1 to T−2 as forecast origins, labels in periods 2 to T−1), a set of m_k fresh non-member lines over the same periods for the membership attack, and two out-of-time test sets of 1,000 fresh lines each at the forecast origin T−1 with labels in period T: a base scenario that follows the realised regional path, and a shock scenario in which an escalating regional shock is imposed at the end of the panel (regional factor set to +0.8 standard deviations at the forecast origin and +1.6 at the horizon, with the features at the origin regenerated accordingly), which represents a regional materials shock, such as the ones reported during the pandemic (Alsharef et al., 2021), arriving after the model has been trained. The population coefficients and the intercept were set so that the training prevalence of disruption is about 15 per cent on average and the Bayes-optimal AUC given the features and the latent states is about 0.90 (realism targets of the protocol as revised after the pilot; the realised values, including the dispersion of prevalence that the shared regional path induces across federations, are reported in Section 4.1). The generator is a form of fully synthetic data in the sense of the disclosure-limitation literature, with parameters set by design rather than estimated from confidential records (Reiter, 2005).

Table 2: Synthetic order-line panel generator: features, blocks, population coefficients, latent processes and heterogeneity parameters

Element	Specification
Delivery performance block	Lead-time deviation (coefficient +0.45), lead-time variability (+0.35), on-time rate (-0.50); within-block correlation 0.40; lead-time deviation and on-time rate correlated -0.45; loadings on supplier state 0.60, 0.40, -0.50 and on regional factor 0.30, 0.20, -0.20
Supplier load block	Supplier backlog (+0.40), capacity utilisation (+0.35), supplier financial stress (+0.30); correlation 0.40; backlog and inventory cover correlated -0.25; loadings on supplier state 0.30 each and on regional factor 0.50, 0.50, 0.20
Order and market block	Order-size ratio (+0.25), three-month price change (+0.30), regional demand index (+0.35, plus a saturating term 0.30 tanh); correlation 0.40; loadings on regional factor 0.20, 0.60, 0.80
Exposure block	Import share (+0.25), transport risk (+0.20), single sourcing (+0.25); correlation 0.40
Buffers and season block	Inventory cover (-0.40, loading -0.30 on regional factor and -0.20 on supplier state), season hazard (+0.20); correlation 0.30
Non-linear terms	Interactions 0.25 capacity utilisation x order size and 0.20 import share x price change; threshold effects +0.50 if on-time rate < -1 SD and +0.40 if inventory cover < -1 SD
Latent processes	Supplier state s: AR(1), persistence 0.70, unit variance, effect 0.80 of next period's state on the logit. Regional factor f: AR(1), persistence 0.60, innovation SD 0.80, shared by all firms, firm loading λ_k, effect 0.55 λ_k of next period's factor on the logit. Unobservable innovations give an unforecastable logit component of SD about 0.72

Element	Specification
Intercept and prevalence	Intercept -3.35; realised prevalence and Bayes-optimal AUC are reported in Table 6
Heterogeneity τ	Per-firm shifts: feature means 0.6 τ SD, coefficients 0.35 τ, intercept 0.5 τ, loading λ_k = 1 + 0.4 τ z (floor 0.2); τ ∈ {0, 0.5, 1}
Panel sizes and periods	m_k lines log-normal with mean 20 and log-SD 0.6, minimum 8; T ∈ {6, 10, 24} periods; training records = m_k (T-2); test 1,000 lines at origin T-1 in base and shock scenarios (regional factor +0.8 at origin, +1.6 at horizon); non-member set m_k lines

Models

Two forecasters are trained. Logistic regression on the fourteen features (15 parameters) is the reference model of rare-event risk scoring (Lessmann et al., 2015; Baesens et al., 2003). A one-hidden-layer neural network with 32 rectified linear units (513 parameters) can represent the interactions and thresholds of the generating process and stands for the flexible models of the construction analytics literature (Pan & Zhang, 2021; Gondia et al., 2020). Both are trained by plain stochastic gradient descent with a batch size of 32, an L2 penalty of 0.0001 and learning rates of 0.1 (logistic regression) and 0.05 (network), the settings of the companion study, which a pilot without privacy confirmed bring the centralised model close to the attainable AUC within 60 epochs. The same optimiser, batch size and number of passes are used in every regime so that the regimes differ only in where the data are and in the privacy mechanism.

Training regimes

Five regimes are compared (Table 3). Local-only training fits each firm's model on its own panel for 60 epochs; it is the status quo. Centralised pooling fits one model on the union of all panels; it is the upper benchmark and is infeasible under the confidentiality constraints. FedAvg runs 60 rounds in which the aggregator broadcasts the global model, every firm runs one local epoch of SGD and returns its update, and the aggregator averages the updates weighted by the number of records (McMahan et al., 2017); secure aggregation is assumed for the transport of updates and changes no number in the utility analysis (Bonawitz et al., 2017). FedAvg with sample-level differential privacy replaces each firm's local SGD by DP-SGD: each record's gradient is clipped to L2 norm C = 1, Gaussian noise with standard deviation σ_k C is added to the sum over a Poisson-sampled batch with sampling rate q_k = 32/n_k, and the sum is divided by the expected batch size (Abadi et al., 2016; Song et al., 2013). FedAvg with client-level differential privacy keeps the local SGD exact and instead clips each firm's whole update to norm 1 and adds Gaussian noise with standard deviation σ_c to the sum at the aggregator before dividing by K; all firms participate in every round and the updates are weighted uniformly, because weighting by size would itself reveal the sizes (McMahan et al., 2018). Sample-level DP bounds what the aggregator and the other firms can learn about any single order-line record of firm k from all of k's updates; client-level DP bounds what can be learned about the presence or the entire panel of one firm. The study measures both because a consortium may need either or both.

Table 3: Training regimes, what each protects, and the hyperparameters used

Regime	Where training happens	What is protected	Hyperparameters
Local only	Each firm on its own panel	Everything (no sharing); no pooled learning	SGD, 60 epochs, batch 32, lr 0.1 (LR) or 0.05 (MLP), L2 0.0001
Centralised pooling	One model on the union of panels (infeasible benchmark)	Nothing	Same optimiser, 60 epochs
FedAvg	60 rounds; one local epoch per round; aggregator averages updates weighted by record count	Raw records stay local; secure aggregation assumed for transport	Same optimiser; all firms every round

Regime	Where training happens	What is protected	Hyperparameters
FedAvg + sample-level DP	As FedAvg with DP-SGD locally: per-record clipping C = 1, Gaussian noise σ_k C, Poisson sampling q_k = 32/n_k	Each line-period record: (ε, 1e-5)-DP per firm over all its steps against aggregator and peers	σ_k calibrated per firm by RDP accounting; ε ∈ {0.5, 1, 2, 4, 8}
FedAvg + client-level DP	As FedAvg; each firm update clipped to norm 1; Gaussian noise σ_c added to the sum at the aggregator; uniform weights	Each firm's whole panel: (ε, 1e-5)-DP over 60 rounds	σ_c calibrated by RDP accounting without subsampling; ε ∈ {4, 16, 64}
Local recalibration (E5)	Platt rescaling of the shared model's scores fitted by each firm on its own records	No exchange; no additional privacy cost against peers	Applied to FedAvg and sample-level DP models
Local fine-tuning (E6)	Five local SGD epochs from the shared model at each firm	No exchange; personalised models	Applied to FedAvg and sample-level DP models

Privacy accounting

The noise multipliers are calibrated with Rényi differential privacy (Mironov, 2017). For sample-level DP, each firm's training is a composition of 60 ceil(n_k/32) sampled Gaussian mechanisms with rate q_k; the Rényi divergence of the sampled Gaussian mechanism is computed with the binomial-expansion bound used by the moments accountant for integer orders (Abadi et al., 2016; Wang et al., 2019); the noise scale is therefore set by the accountant rather than by the analytic calibration of a single Gaussian release (Balle & Wang, 2018), summed over the steps, and converted to (ε, δ) at δ = 1e-5 with the standard conversion (Mironov, 2017). The smallest σ_k whose accounted ε is at most the target is found by bisection for every firm, so that firms with fewer records, which have higher sampling rates and fewer steps, receive the noise their own budget requires. For client-level DP with all firms in every round there is no subsampling, the per-round divergence is $\alpha/(2\sigma_c^2)$, and σ_c is calibrated over the 60 rounds in the same way. The accountant is the one of the companion study, checked against two published values (ε = 1.26 for q = 0.01, σ = 4 and 10,000 steps at δ = 1e-5 (Abadi et al., 2016); ε = 3.0 for σ = 1.1, q = 256/60,000 and 60 epochs). Because a record here is a line-period and records within a line are correlated, the sample-level guarantee protects a single line-period, not the whole history of a line; group privacy would multiply ε by the number of periods for a guarantee at line level, a limitation returned to in Section 6. The targets are ε ∈ {0.5, 1, 2, 4, 8} for sample-level DP and ε ∈ {4, 16, 64} for client-level DP.

Evaluation

Table 4 lists the measures. Utility is measured out of time on each firm's 1,000 test lines at the forecast origin and pooled over firms, in the base scenario and in the shock scenario. Discrimination is the area under the receiver operating characteristic curve (AUC), pooled, as the mean over firms and as the minimum over firms (Hanley & McNeil, 1982; Bradley, 1997); the Bayes-optimal AUC given the features and the latent states is reported as the attainable ceiling. Calibration is the Brier score (Brier, 1950). Top-decile capture is the share of all next-period disruptions among the ten per cent of lines with the highest scores, which corresponds to a watch list of fixed size. Decision quality is the expected misclassification cost per line when lines are flagged for buffering or alternative sourcing at the cost-optimal threshold for calibrated probabilities (Elkan, 2001), with a missed disruption costing eight times a false alarm as the headline ratio (threshold 1/9) and four and sixteen times as alternatives; the cost is compared with the best naive policy (flag nothing or flag everything) and reported as the relative saving, which is positive when the forecast is worth using. The shock penalty is the difference between the base and the shock scenario in AUC and in relative saving for the same model. Per-firm AUC is retained to measure the dispersion of utility and its relation to firm size (Bagdasaryan et al., 2019).

Empirical privacy is measured with a loss-threshold membership-inference attack against the released model: for each firm, the per-record loss of the model is computed on its training records (members) and on the records of fresh lines of the same firm over the same periods (non-members), and an attacker who labels a record as a member when its loss is below a threshold is evaluated by the best advantage

over thresholds and by the attack AUC (Yeom et al., 2018; Salem et al., 2019; Jayaraman & Evans, 2019). The best-threshold advantage favours the attacker and is an upper bound for this attack family; the generalisation gap is recorded because the attack's success is driven by it (Yeom et al., 2018). In the local regime the attack is run against each firm's own model; in the other regimes against the shared model.

Table 4: Evaluation measures: utility in base and shock scenarios, dispersion and empirical privacy

Measure	Definition	Role
AUC (pooled, mean, minimum)	Area under the ROC curve on 1,000 out-of-time test lines per firm at the forecast origin: pooled over firms; mean and minimum of per-firm values; base and shock scenarios	Discrimination; dispersion across firms
Attainable AUC	AUC of the true next-period probability given the features and the latent states at the origin	Ceiling set by the unforecastable innovations
Brier score	Mean squared difference between predicted probability and outcome	Calibration and refinement
Top-decile capture	Share of all next-period disruptions among the 10% of lines with the highest scores	Watch list of fixed size
Expected cost and relative saving	Cost per line at the threshold 1/9 (missed disruption = 8, false alarm = 1; also 4 and 16) and its saving relative to the best naive policy	Decision value for buffering and alternative sourcing
Shock penalty	Base-scenario value minus shock-scenario value of AUC and of relative saving for the same model	Robustness to a regional shock at the horizon
Per-firm AUC difference	Difference between a firm's AUC under a regime and under FedAvg on the same seed, related to its record count	Who pays for privacy
Membership-inference advantage and AUC	Loss-threshold attack: best TPR minus FPR over thresholds, and attack AUC, members vs fresh non-member lines, per firm	Empirical privacy
Generalisation gap	Mean loss on non-members minus mean loss on members	Driver of the attack

Experimental design and analysis

Table 5 lists the experiments. E1 is the main factorial: both models, three heterogeneity levels, eleven regimes, K = 10 firms, 20 lines on average over T = 10 periods, ten seeds. E2 varies the number of firms (5, 10, 20, 50, 100) at $\tau = 0.5$ with eight regimes. E3 varies the length of history (6, 10 and 24 periods, that is, 4, 8 and 22 forecast origins per line). E4 is a sensitivity analysis of the sample-level mechanism at $\varepsilon = 2$: clipping norm (0.5, 1, 2, 4), local epochs (1, 2, 5) and rounds (30, 60, 120). E5 adds local post-hoc recalibration, a monotone two-parameter logistic rescaling of the shared model's scores fitted by each firm on its own training records, which involves no exchange (Guo et al., 2017); when the unconstrained fit returns a non-positive slope only the intercept is refitted. E6 adds local fine-tuning, five epochs of local SGD from the shared model at each firm (Fallah et al., 2020; Smith et al., 2017). Each seed generates a new federation, a new regional path and new initialisations; the same seed is used across regimes so that comparisons are paired. In all, 1,840 training runs were executed (1.1 processor-hours). Results are reported as means over seeds with 95 per cent t-intervals; differences between regimes are paired by seed and tested with paired t-tests, with the reservation that ten seeds give limited power and that the many comparisons are descriptive rather than confirmatory (Demšar, 2006). All code, the generator, the accountant and every run's record are in the supplementary archive, and every number in the text and tables is produced from the run records by the analysis script.

Table 5: Experimental design: experiments, factors, levels, regimes and seeds

Experiment	Factors and levels	Regimes	Models	Seeds	Runs
E1 main factorial	τ ∈ {0, 0.5, 1}; K = 10; 20 lines; T = 10	Local, central, FedAvg, sample-level DP (ε 0.5, 1, 2, 4, 8), client-level DP (ε 4, 16, 64)	LR, MLP	10	660
E2 federation size	K ∈ {5, 20, 50, 100}; τ = 0.5; T = 10	Local, central, FedAvg, sample-level DP (ε 1, 4), client-level DP (ε 4, 16, 64)	LR, MLP	10	640
E3 history length	T ∈ {6, 24}; τ = 0.5; K = 10	Local, central, FedAvg, sample-level DP (ε 1, 4)	LR, MLP	10	200
E4 sensitivity	C ∈ {0.5, 1, 2, 4} x local epochs ∈ {1, 2, 5} at ε = 2; rounds ∈ {30, 120}	Sample-level DP (ε 2); FedAvg for the rounds factor	LR	5	80
E5 recalibration	τ = 0.5; K = 10; T = 10	FedAvg and sample-level DP (ε 1, 2, 4) with local Platt rescaling	LR, MLP	10	80
E6 fine-tuning	τ ∈ {0, 0.5, 1}; K = 10; T = 10	FedAvg and sample-level DP (ε 1, 4) with five local epochs	LR, MLP	10	180

RESULTS

Realised federations and privacy calibration

Table 6 describes the simulated federations. With a mean of 20 lines over 10 periods, the firms held between 8 and 76 material lines (median 14.5) and between 64 and 608 training records (median 116), so that a typical federation combined a few large buyers with many small ones. The disruption prevalence in the training records was 16.6 per cent on average but varied widely across federations (standard deviation 8.1, range 6 to 38 per cent), because every firm in a federation shares one regional path and a boom or a slump in that path moves the base rate of the whole panel; in the test period the prevalence was 27.0 per cent in the base scenario and 35.8 per cent under the imposed shock. The imposed path (+0.8 at the origin, +1.6 at the horizon) is fixed whereas the realised path varies: in four of the ten federations of the main cell the realised factor at the origin already exceeded +0.8 (the realised origin values ranged from -0.73 to 1.75 and the base test prevalence from 3 to 57 per cent), so that for those federations the shock scenario lowered rather than raised the base rate. The shock comparison is therefore between the realised path and one fixed adverse path, not between calm and crisis in every federation, and Section 4.5 also reports the subset in which the imposed path did raise the base rate. The attainable AUC given the features and the latent states was 0.905 at τ = 0.5 in the base scenario and 0.894 under the shock, within the realism targets. Table 7 gives the noise multipliers that the accountant assigned. At n_k = 112 records (sampling rate 0.29, 240 steps) the sample-level guarantee of ε = 1 requires σ = 21.82 and ε = 4 requires σ = 5.87; a firm with 56 records needs σ = 30.79 at ε = 1 and one with 448 records σ = 10.25, because fewer records mean a higher sampling rate and less amplification by subsampling, and the multipliers are larger throughout than for the firms of the companion study because the panels are smaller. For client-level DP over 60 rounds, ε = 4 requires σ_c = 10.05, ε = 16 requires σ_c = 2.97 and ε = 64 requires σ_c = 1.07 on updates clipped to norm 1; divided by K = 10 firms this is noise of standard deviation 1.01 and 0.3 per coordinate of the average update, comparable to or larger than the averaged update itself.

Table 6: Realised characteristics of the simulated federations and the computational budget

Characteristic	Value
Lines per firm m_k (E1)	8 to 76; median 14.5
Training records per firm n_k (E1, T = 10)	64 to 608; median 116
Disruption prevalence in training records (E1)	16.6% on average (SD 8.1 across federations; range 6% to 38%)
Prevalence in the test period, base scenario	27.0% (SD 17.5)

Characteristic	Value
Prevalence in the test period, shock scenario	35.8% (SD 3.3)
Attainable (Bayes) AUC, τ = 0: base / shock	0.904 (SD 0.012) / 0.890 (SD 0.003)
Attainable (Bayes) AUC, τ = 0.5: base / shock	0.905 (SD 0.014) / 0.894 (SD 0.006)
Attainable (Bayes) AUC, τ = 1: base / shock	0.914 (SD 0.014) / 0.904 (SD 0.008)
Runs executed	1,840 (1.1 processor-hours)

Table 7: Noise multipliers σ calibrated by Rényi-DP accounting for the sample-level mechanism, by firm size and budget (60 rounds, batch 32, δ = 1e-5)

Records n_k	Sampling rate q	Steps	ε = 0.5	ε = 1	ε = 2	ε = 4	ε = 8
56	0.571	120	60.85	30.79	15.75	8.21	4.43
112	0.286	240	43.06	21.82	11.19	5.87	3.21
224	0.143	420	28.51	14.46	7.45	3.94	2.2
448	0.071	840	20.18	10.25	5.29	2.84	1.64

Value of federation without privacy noise (RQ1)

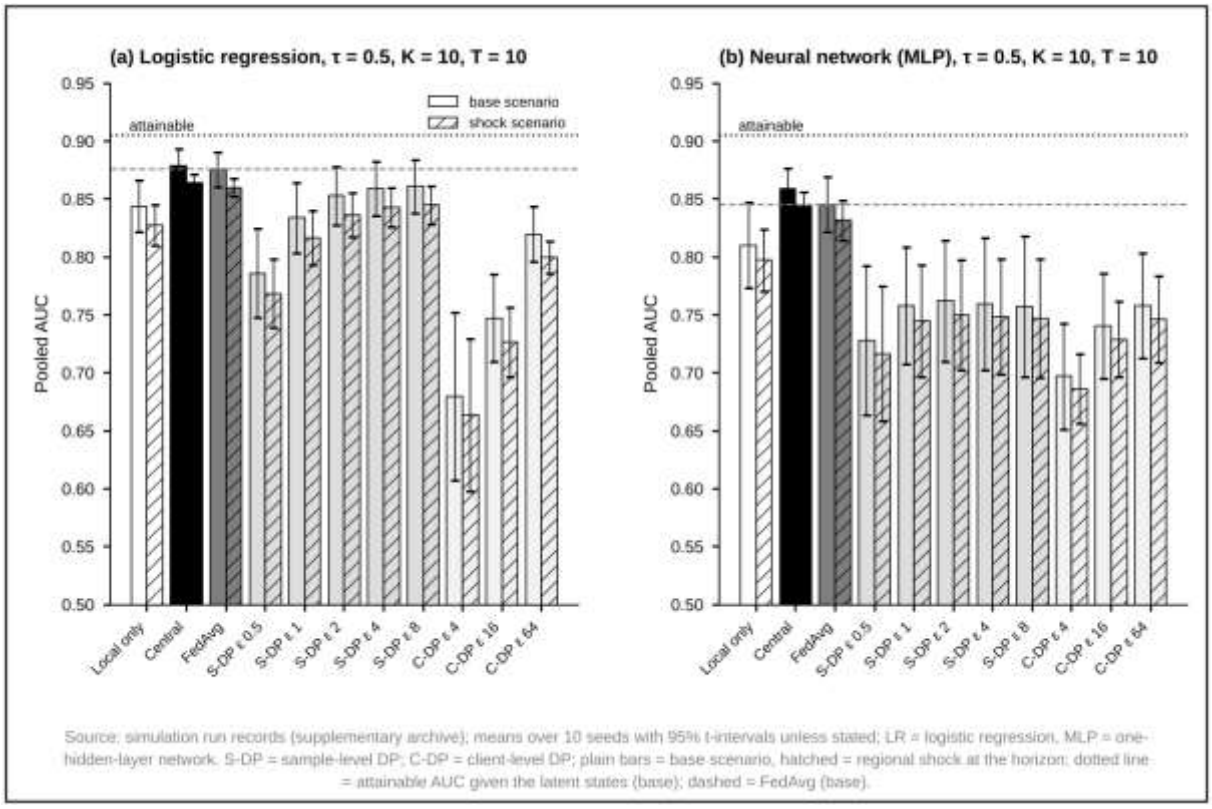
Table 8 and Figure 3 report the main results at moderate heterogeneity (τ = 0.5) with ten firms in the base scenario. For logistic regression, local-only forecasters reached a pooled AUC of 0.844 with a mean per-firm AUC of 0.844 and a minimum of 0.780; centralised pooling reached 0.879 (minimum 0.842) and FedAvg 0.876 (95% CI 0.861 to 0.891; minimum 0.837), against an attainable 0.905. Federation recovered almost all of the pooling benefit: FedAvg was +3.2 points (95% CI +2.2 to +4.2) above local training in mean per-firm AUC ($p < 0.001$) and -0.3 points (95% CI -0.7 to +0.0) from the centralised benchmark ($p = 0.071$), and the worst-served firm improved from an AUC of 0.780 to 0.837. For the neural network, local training was weaker (0.810 pooled; minimum 0.711), FedAvg reached 0.845 and centralised training 0.859; the network gained more from federation than logistic regression (+3.6 points (95% CI +2.2 to +5.0) in mean per-firm AUC) but fell further short of pooling (-1.4 points (95% CI -2.3 to -0.6)), a pattern consistent with the client drift that heterogeneous panels induce in federated averaging (Li et al., 2020d; Hsieh et al., 2020), although the simulation does not isolate the cause. The decision metrics followed discrimination: the logistic FedAvg forecaster captured 37 per cent of all next-period disruptions in the riskiest tenth of lines and saved 33% of the expected cost of the best naive policy at the 8:1 cost ratio, against 26% for local forecasters and 34% for pooling; at 4:1 the savings were 35% and 26%, and at 16:1 25% and 21%. The local networks were not worth using for decisions (1%), whereas the federated network saved 24%.

Table 8: Main results at τ = 0.5, K = 10, T = 10, base scenario: discrimination, dispersion, calibration, watch-list capture, decision value (8:1) and membership-inference advantage by model and regime

Model and regime	AUC pooled (95% CI)	AUC mean	AUC min	Brier	Top-decile capture	Cost saving	MIA advantage
LR: Local only	0.844 (0.821 to 0.866)	0.844	0.780	0.122	34%	26%	0.095
LR: Centralised pooling	0.879 (0.865 to 0.893)	0.879	0.842	0.110	37%	34%	0.073
LR: FedAvg	0.876 (0.861 to 0.891)	0.875	0.837	0.109	37%	33%	0.074

Model and regime	AUC pooled (95% CI)	AUC mean	AUC min	Brier	Top-decile capture	Cost saving	MIA advantage
LR: Sample-level DP, ε = 0.5	0.786 (0.747 to 0.824)	0.789	0.739	0.167	29%	6%	0.076
LR: Sample-level DP, ε = 1	0.834 (0.804 to 0.864)	0.835	0.792	0.132	33%	24%	0.072
LR: Sample-level DP, ε = 2	0.853 (0.828 to 0.878)	0.854	0.812	0.121	35%	30%	0.071
LR: Sample-level DP, ε = 4	0.859 (0.835 to 0.883)	0.859	0.819	0.118	36%	31%	0.071
LR: Sample-level DP, ε = 8	0.861 (0.838 to 0.884)	0.861	0.821	0.118	36%	32%	0.072
LR: Client-level DP, ε = 4	0.679 (0.607 to 0.752)	0.682	0.628	0.352	21%	-47%	0.078
LR: Client-level DP, ε = 16	0.747 (0.709 to 0.785)	0.748	0.692	0.232	26%	-11%	0.075
LR: Client-level DP, ε = 64	0.820 (0.796 to 0.844)	0.819	0.770	0.146	32%	19%	0.072
MLP: Local only	0.810 (0.773 to 0.847)	0.808	0.711	0.139	32%	1%	0.119
MLP: Centralised pooling	0.859 (0.841 to 0.876)	0.858	0.825	0.120	36%	25%	0.082
MLP: FedAvg	0.845 (0.821 to 0.869)	0.844	0.807	0.126	35%	24%	0.078
MLP: Sample-level DP, ε = 0.5	0.728 (0.664 to 0.793)	0.729	0.674	0.179	25%	-73%	0.070
MLP: Sample-level DP, ε = 1	0.758 (0.707 to 0.809)	0.759	0.712	0.182	27%	-68%	0.073
MLP: Sample-level DP, ε = 2	0.762 (0.710 to 0.814)	0.762	0.718	0.189	28%	-79%	0.075
MLP: Sample-level DP, ε = 4	0.760 (0.703 to 0.817)	0.759	0.715	0.194	28%	-92%	0.074
MLP: Sample-level DP, ε = 8	0.757 (0.697 to 0.818)	0.756	0.711	0.196	28%	-100%	0.074
MLP: Client-level DP, ε = 4	0.697 (0.651 to 0.743)	0.696	0.640	0.280	22%	-118%	0.250
MLP: Client-level DP, ε = 16	0.740 (0.695 to 0.786)	0.740	0.685	0.244	26%	-93%	0.081
MLP: Client-level DP, ε = 64	0.758 (0.712 to 0.804)	0.757	0.706	0.197	27%	-53%	0.080

Figure 3: Pooled AUC by regime for logistic regression and the neural network at τ = 0.5, K = 10, T = 10, in the base and the shock scenario, with the attainable AUC and the FedAvg reference



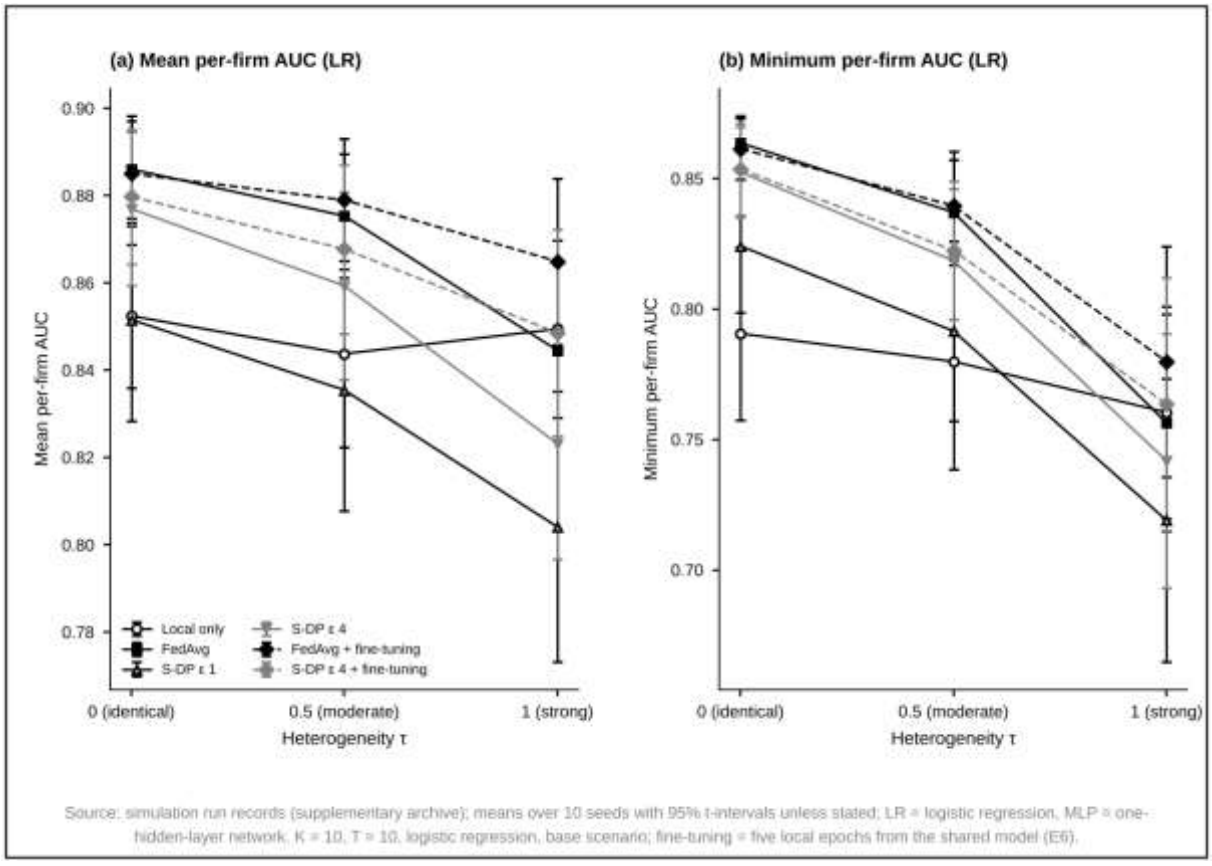
Heterogeneity changed the value of a shared forecaster (Table 9 and Figure 4). With identically distributed firms (τ = 0), FedAvg matched centralised pooling for logistic regression (0.886 against 0.886) and was +3.4 points (95% CI +2.7 to +4.0) above local forecasters. With strong heterogeneity (τ = 1), one shared forecaster was no longer the best choice for every firm: local forecasters reached a mean per-firm AUC of 0.849 against 0.845 for FedAvg (-0.5 points (95% CI -1.4 to +0.4), p = 0.259) and 0.847 for pooling, while the minimum per-firm AUC was similar (0.756 for FedAvg against 0.760 for local training). The pattern is the one the heterogeneity literature predicts: a single model averaged across firms with different material mixes, exposures and risk relationships fits firms with distinct risk relationships less well than their own models (Li et al., 2020c; Smith et al., 2017; Mohri et al., 2019). Local fine-tuning of the shared forecaster (E6) resolved the conflict at every level of heterogeneity: five local epochs from the FedAvg model gave a mean per-firm AUC of 0.865 at τ = 1, above local training (+1.5 points (95% CI +0.9 to +2.2), p < 0.001) and above FedAvg (+2.0 points (95% CI +1.7 to +2.4)), with a minimum of 0.780, and it also improved on FedAvg at τ = 0.5 (0.879 against 0.875) and cost a tenth of a point at τ = 0 (0.885 against 0.886).

Table 9: Mean and minimum per-firm AUC by heterogeneity level and regime (K = 10, T = 10, base scenario), including local fine-tuning

Model and regime	τ = 0 (identical)	τ = 0.5 (moderate)	τ = 1 (strong)
LR: Local only	0.852 / 0.791	0.844 / 0.780	0.849 / 0.760
LR: Centralised pooling	0.886 / 0.864	0.879 / 0.842	0.847 / 0.759
LR: FedAvg	0.886 / 0.864	0.875 / 0.837	0.845 / 0.756
LR: Sample-level DP, ε = 1	0.852 / 0.824	0.835 / 0.792	0.804 / 0.719
LR: Sample-level DP, ε = 4	0.877 / 0.852	0.859 / 0.819	0.823 / 0.742
LR: FedAvg + local fine-tuning	0.885 / 0.861	0.879 / 0.840	0.865 / 0.780

Model and regime	τ = 0 (identical)	τ = 0.5 (moderate)	τ = 1 (strong)
LR: Sample-level DP ε = 4 + fine-tuning	0.880 / 0.854	0.868 / 0.822	0.849 / 0.764
MLP: Local only	0.808 / 0.710	0.808 / 0.711	0.820 / 0.707
MLP: Centralised pooling	0.868 / 0.842	0.858 / 0.825	0.832 / 0.747
MLP: FedAvg	0.854 / 0.828	0.844 / 0.807	0.816 / 0.728
MLP: Sample-level DP, ε = 1	0.766 / 0.735	0.759 / 0.712	0.737 / 0.659
MLP: Sample-level DP, ε = 4	0.768 / 0.738	0.759 / 0.715	0.738 / 0.648
MLP: FedAvg + local fine-tuning	0.854 / 0.824	0.847 / 0.806	0.832 / 0.748
MLP: Sample-level DP ε = 4 + fine-tuning	0.813 / 0.774	0.803 / 0.750	0.794 / 0.704

Figure 4: Mean and minimum per-firm AUC against heterogeneity for local, federated, sample-level DP and fine-tuned forecasters (logistic regression)



Privacy-utility curves (RQ2)

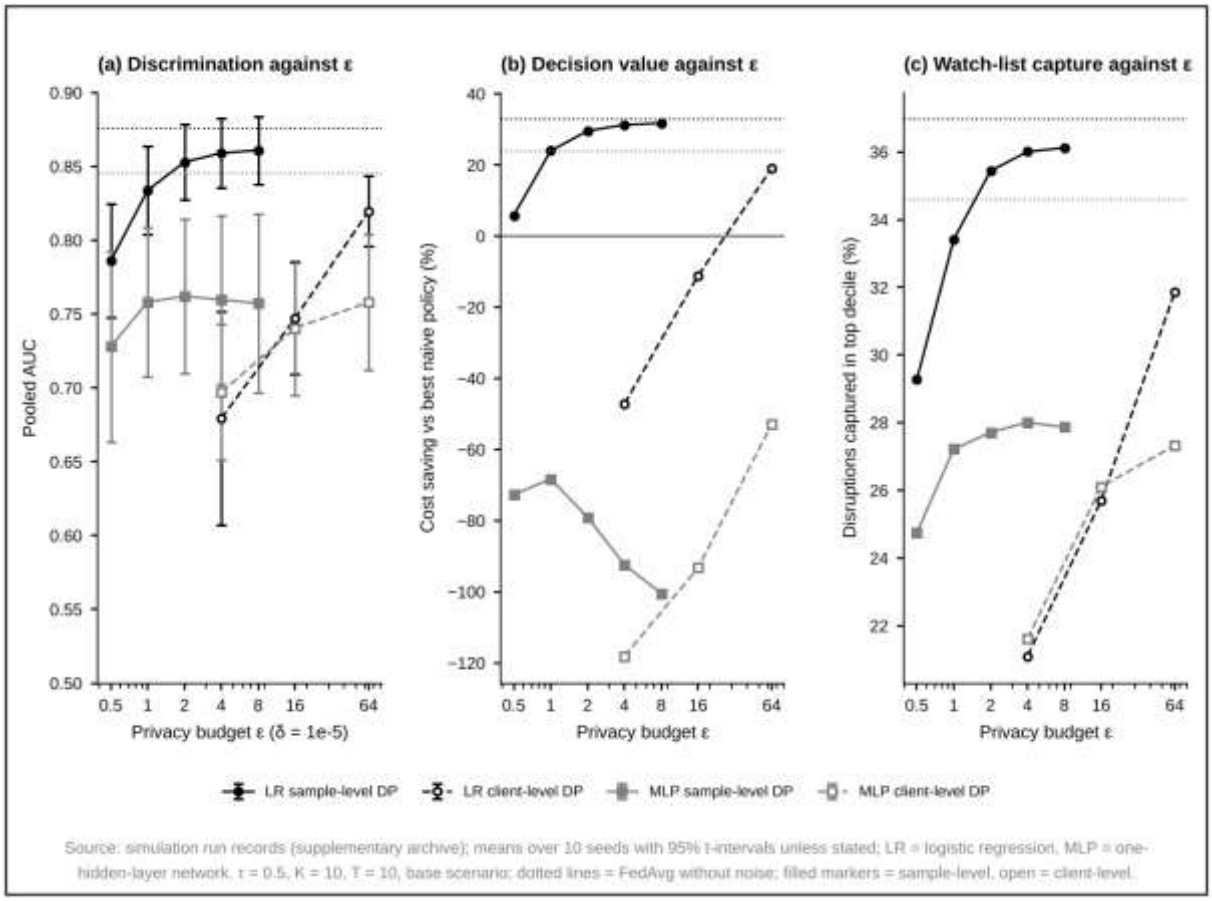
Table 10 and Figure 5 give the paired differences against FedAvg on the same seeds. Sample-level differential privacy was affordable for logistic regression but cost more than in a setting with larger data sets, because the panels are small and the multipliers of Table 7 correspondingly large. At ε = 8 the pooled AUC fell by 1.5 points (p = 0.013), at ε = 4 by 1.7 points (paired difference -1.7 points (95% CI -2.9 to -0.5)), at ε = 2 by 2.3 points, at ε = 1 by 4.2 points and at ε = 0.5 by 9.0 points. Down to ε = 2 the private federated forecaster was still ahead of local training in pooled AUC (0.853 against 0.844) and for the worst-served firm (minimum 0.812 against 0.780), although neither paired difference reached significance (p = 0.128 and p = 0.074); at ε = 1 it fell to the level of local training in pooled AUC (0.834 against 0.844) and no worse for the worst-served firm (0.792 against 0.780, a difference well within its interval); at ε = 0.5 it was worse than local training on both counts. The decision metrics degraded faster

than discrimination. The relative cost saving at 8:1 fell from 33% for FedAvg to 31% at ε = 4, 24% at ε = 1 and 6% at ε = 0.5, and the Brier score rose from 0.109 to 0.132 at ε = 1: the noisy gradients shrink the coefficients towards zero, the forecasts become under-confident, fewer lines cross the cost-optimal threshold and more disruptions are missed. Top-decile capture, which depends on ranking only, fell from 37 to 33 per cent at ε = 1.

Table 10: Paired differences against FedAvg on the same seeds, all multiplied by 100 (AUC and MIA advantage in points, top-decile capture in percentage points, Brier score and expected cost per line in hundredths) at τ = 0.5, K = 10, T = 10, base scenario; asterisk: paired t-test p < 0.05

Model and regime	ΔAUC pooled	ΔAUC mean	ΔAUC min	ΔTop-decile	ΔBrier	ΔCost	ΔMIA
LR: Local only	-3.2*	-3.2*	-5.7*	-2.7*	+1.3*	+3.6*	+2.1*
LR: Centralised pooling	+0.3	+0.3	+0.5*	+0.2	+0.1	-1.0	-0.1
LR: Sample-level DP, ε = 0.5	-9.0*	-8.6*	-9.8*	-7.7*	+5.8*	+15.7*	+0.2
LR: Sample-level DP, ε = 1	-4.2*	-4.0*	-4.5*	-3.5*	+2.3*	+5.4*	-0.2
LR: Sample-level DP, ε = 2	-2.3*	-2.2*	-2.5*	-1.5*	+1.3*	+2.1*	-0.3
LR: Sample-level DP, ε = 4	-1.7*	-1.6*	-1.8*	-1.0*	+1.0*	+1.1*	-0.3
LR: Sample-level DP, ε = 8	-1.5*	-1.4*	-1.6*	-0.8*	+0.9*	+0.8	-0.2
LR: Client-level DP, ε = 4	-19.6*	-19.3*	-20.9*	-15.9*	+24.4*	+48.6*	+0.4
LR: Client-level DP, ε = 16	-12.9*	-12.7*	-14.5*	-11.3*	+12.3*	+26.5*	+0.1
LR: Client-level DP, ε = 64	-5.6*	-5.6*	-6.7*	-5.1*	+3.7*	+7.7*	-0.2
MLP: Local only	-3.5*	-3.6*	-9.6*	-2.2*	+1.3*	+11.9*	+4.1*
MLP: Centralised pooling	+1.4*	+1.4*	+1.9*	+1.0	-0.6	-0.8	+0.4
MLP: Sample-level DP, ε = 0.5	-11.7*	-11.5*	-13.3*	-9.9*	+5.3*	+59.7*	-0.9
MLP: Sample-level DP, ε = 1	-8.7*	-8.5*	-9.5*	-7.4*	+5.6*	+55.5*	-0.5
MLP: Sample-level DP, ε = 2	-8.3*	-8.2*	-8.9*	-6.9*	+6.3*	+59.2*	-0.3
MLP: Sample-level DP, ε = 4	-8.6*	-8.5*	-9.2*	-6.6*	+6.8*	+65.0*	-0.4
MLP: Sample-level DP, ε = 8	-8.8*	-8.8*	-9.6*	-6.7*	+7.0*	+68.6*	-0.5
MLP: Client-level DP, ε = 4	-14.8*	-14.8*	-16.7*	-13.0*	+15.4*	+82.1*	+17.1*
MLP: Client-level DP, ε = 16	-10.5*	-10.4*	-12.2*	-8.5*	+11.8*	+65.7*	+0.3
MLP: Client-level DP, ε = 64	-8.7*	-8.7*	-10.1*	-7.3*	+7.1*	+42.4*	+0.2

Figure 5: Privacy-utility curves at τ = 0.5, K = 10, T = 10: pooled AUC, cost saving and top-decile capture against ε for sample-level and client-level differential privacy, both models



The neural network paid far more. Under sample-level DP its pooled AUC fell by 8.8 points at ε = 8 and 8.7 points at ε = 1, to 0.760 at ε = 4, below the local logistic forecasters, and its cost saving was negative at every budget (-92% at ε = 4), that is, the private network was worth less than the naive policy. The reason is dimensional: with 513 parameters, clipping to norm 1 and the multipliers of Table 7, the noise added per step is spread over many more coordinates than for the 15-parameter logistic model, and the same budget buys much less signal (Bassily et al., 2014; Chaudhuri et al., 2011; Jayaraman & Evans, 2019). Under privacy noise and with about a hundred records per firm, the low-dimensional forecaster is the better forecaster.

Client-level differential privacy was not affordable with ten firms. For logistic regression the pooled AUC fell by 19.6 points at ε = 4 (to 0.679), by 12.9 points at ε = 16 and by 5.6 points at ε = 64; the cost saving was negative at ε = 4 and ε = 16 (-11%) and reached 19% only at ε = 64, a budget that offers little formal protection. The mechanism is working as its analysis predicts: the noise that protects one firm's whole contribution must be comparable to that contribution, and it is divided only by the number of participants (McMahan et al., 2018; Wei et al., 2020).

Federation size and history length (RQ1 continued)

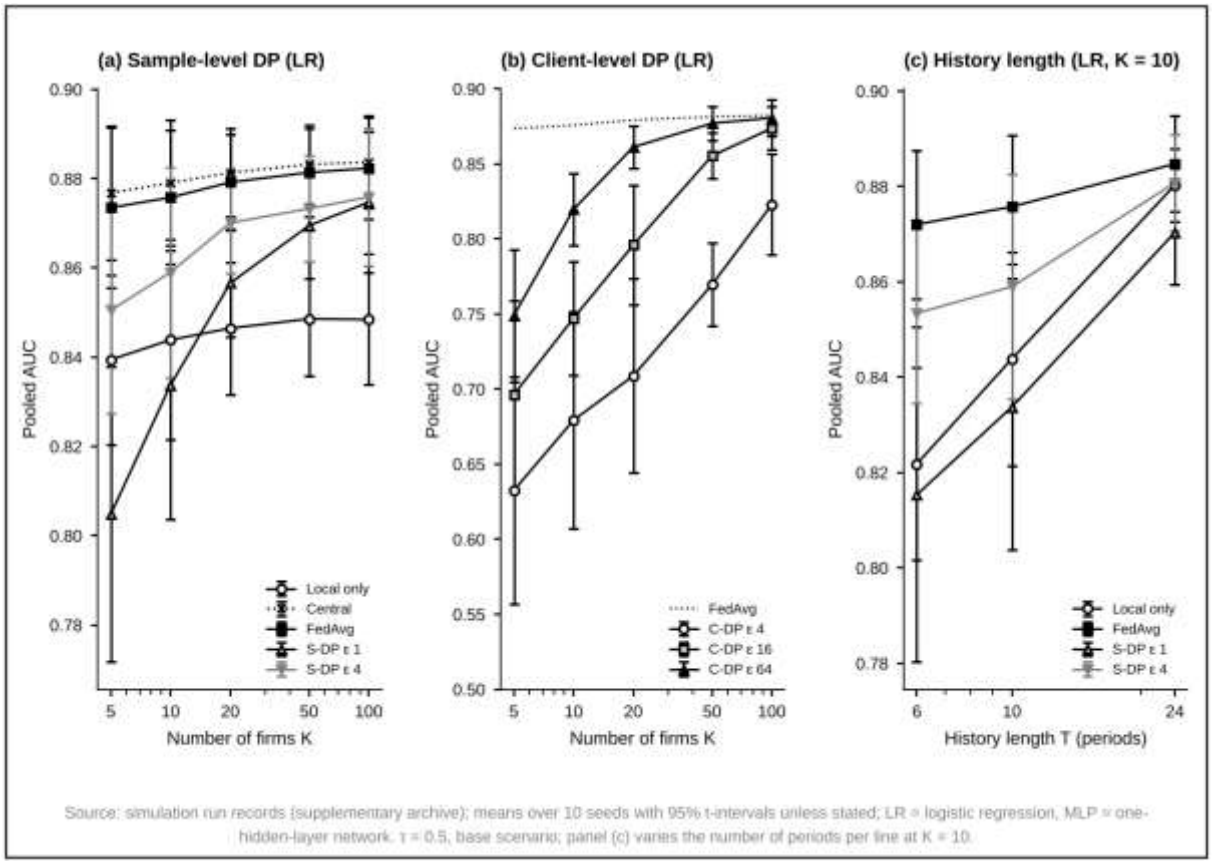
Table 11 and Figure 6 vary the number of firms at τ = 0.5. Without privacy noise the value of federation was present with five firms (FedAvg 0.873 against local 0.839) and changed little with size: the gap between FedAvg and local training was between 3.2 and 3.4 points across the five sizes, consistent with a pooled total of several hundred records being already ample for a 15-parameter model at K = 5. The sample-level mechanism improved with K because the global model averages the noise of more firms: at ε = 1 the loss against FedAvg was 6.9 points with K = 5, 4.2 with K = 10, 2.3 with K = 20, 1.2 with K = 50 and 0.8 with K = 100. The client-level mechanism improved faster but from a much lower base: at ε = 16 the loss was 17.7 points with K = 5, 12.9 with K = 10, 2.6 with K = 50 and 0.9 with K = 100; at ε = 4 it was still 5.9 points with 100 firms. A client-level guarantee at a conventional budget is therefore out

of reach for consortia of the sizes that exist in regional construction markets.

Table 11: Pooled AUC by number of firms K and regime (τ = 0.5, T = 10, base scenario)

Model and regime	K = 5	K = 10	K = 20	K = 50	K = 100
LR: Local only	0.839	0.844	0.846	0.849	0.848
LR: Centralised pooling	0.877	0.879	0.881	0.883	0.884
LR: FedAvg	0.873	0.876	0.879	0.881	0.882
LR: Sample-level DP, ε = 1	0.805	0.834	0.857	0.869	0.875
LR: Sample-level DP, ε = 4	0.851	0.859	0.870	0.873	0.876
LR: Client-level DP, ε = 4	0.633	0.679	0.709	0.770	0.823
LR: Client-level DP, ε = 16	0.696	0.747	0.796	0.855	0.874
LR: Client-level DP, ε = 64	0.749	0.820	0.861	0.877	0.881
MLP: Local only	0.808	0.810	0.817	0.819	0.819
MLP: Centralised pooling	0.850	0.859	0.871	0.879	0.882
MLP: FedAvg	0.843	0.845	0.850	0.856	0.857
MLP: Sample-level DP, ε = 1	0.723	0.758	0.759	0.762	0.771
MLP: Sample-level DP, ε = 4	0.755	0.760	0.771	0.773	0.777
MLP: Client-level DP, ε = 4	0.589	0.697	0.707	0.744	0.763
MLP: Client-level DP, ε = 16	0.723	0.740	0.727	0.799	0.835
MLP: Client-level DP, ε = 64	0.739	0.758	0.788	0.841	0.850

Figure 6: Pooled AUC against the number of firms for the sample-level and client-level mechanisms, and against the length of history (logistic regression, τ = 0.5)



Source: simulation run records (supplementary archive); means over 10 seeds with 95% t-intervals unless stated; LR = logistic regression, MLP = one-hidden-layer network. τ = 0.5, base scenario; panel (c) varies the number of periods per firm at K = 10.

The length of history mattered as much as the budget (Table 12 and Figure 6c). With six periods (four forecast origins per line, a median of 74 records per firm), local logistic forecasters were 5.0 points below FedAvg and sample-level DP at ε = 1 cost 5.7 points; with 24 periods (median 330 records), local forecasters were within 0.4 points of FedAvg and ε = 1 cost 1.4 points. Federation is most valuable, and the privacy budget most expensive, precisely for the firms with the short histories that construction's project-based procurement produces: with six periods the noise-free federated forecaster gained 5.0 points over local training, ε = 4 kept most of that gain (0.853 against 0.822 for local training), and at ε = 1 the private forecaster was level with local training (0.815 against 0.822).

Table 12: Pooled and minimum per-firm AUC by history length T and regime (τ = 0.5, K = 10, base scenario)

Model and regime	T = 6 periods (pooled / min)	T = 10	T = 24
LR: Local only	0.822 / 0.719	0.844 / 0.780	0.880 / 0.831
LR: Centralised pooling	0.874 / 0.827	0.879 / 0.842	0.885 / 0.843
LR: FedAvg	0.872 / 0.825	0.876 / 0.837	0.885 / 0.844
LR: Sample-level DP, ε = 1	0.815 / 0.755	0.834 / 0.792	0.870 / 0.826
LR: Sample-level DP, ε = 4	0.853 / 0.803	0.859 / 0.819	0.881 / 0.838
MLP: Local only	0.782 / 0.650	0.810 / 0.711	0.859 / 0.802
MLP: Centralised pooling	0.851 / 0.808	0.859 / 0.825	0.878 / 0.837
MLP: FedAvg	0.834 / 0.787	0.845 / 0.807	0.870 / 0.828
MLP: Sample-level DP, ε = 1	0.713 / 0.656	0.758 / 0.712	0.823 / 0.781
MLP: Sample-level DP, ε = 4	0.729 / 0.673	0.760 / 0.715	0.837 / 0.794

Who pays, and what a regional shock does (RQ3)

Figure 7 shows who paid within a federation. Relating each firm's AUC under a regime to its AUC under FedAvg on the same seed, the benefit of federation over local training was largest for the smallest third of firms (local forecasters 4.6 points below FedAvg) and smallest for the largest third (1.6 points), with a positive correlation between size and the local-minus-FedAvg difference (Spearman ρ = 0.51, n = 100 firm-seeds). Under sample-level DP the loss against FedAvg was spread more evenly across sizes than the per-firm calibration of σ might suggest: at ε = 1 the smallest third lost 3.3 points, the middle third 5.2 and the largest third 3.7 (ρ = -0.08), a loss unrelated to size, because the shared forecaster is the same for everyone and the noise of a small firm, weighted by its size in the average, is diluted by the others. The disparate impact reported for differential privacy on under-represented groups (Bagdasaryan et al., 2019) appears here mainly in the comparison of local and federated training, not within the private federated forecaster.

Figure 7: Per-firm AUC difference against FedAvg plotted against firm size, and mean difference by size tercile (logistic regression, τ = 0.5)

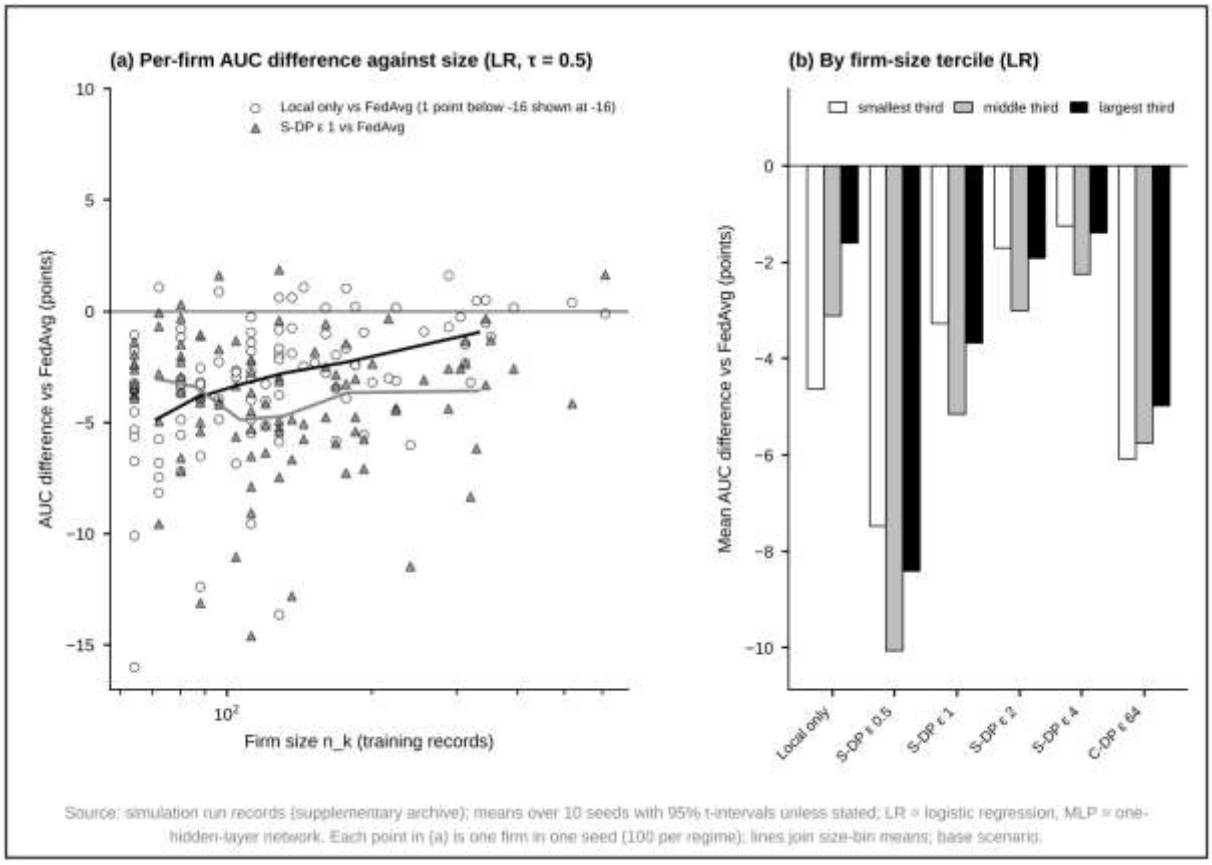


Table 13 and the hatched bars of Figure 3 report what happened when a regional shock arrived at the forecast horizon. Discrimination was robust: the shock lowered the pooled AUC of the FedAvg logistic forecaster from 0.876 to 0.860, of the local forecasters from 0.844 to 0.827 and of the ε = 1 private forecaster from 0.834 to 0.817, a penalty of about 1.6 to 1.7 points of the same order as the fall in the attainable AUC itself (0.905 to 0.894), because the ranking of lines by their period-t features survives a common shift. Decision value did not: the base rate rose from 27 to 36 per cent, forecasts calibrated on calmer periods under-predicted the risk of every line, fewer lines crossed the expected-cost threshold, and the relative saving at 8:1 fell from 33% to 24% for FedAvg, from 26% to 13% for the local forecasters, from 34% to 18% for pooling and from 24% to 8% at ε = 1; the Brier score of the FedAvg forecaster rose from 0.109 to 0.148. Under the shock the federated forecaster kept more of its decision value than the isolated local ones (24% against 13%; paired difference in saving -11 points (95% CI -15 to -7) per cent, $p < 0.001$), which is the benefit of having learned the response to the regional indicators from every member's history rather than from a single firm's few months. Restricted to the six federations in which the imposed path lay above the realised one, so that the base rate rose (from 14 to 36 per cent), the pattern was the same and sharper: the FedAvg saving fell from 44% to 25%, the local saving from 38% to 14%, the pooled saving from 47% to 20% and the ε = 1 saving from 37% to 9%, while the pooled AUC of the FedAvg forecaster fell by 2.4 points. The saving lost under the shock was 8.5 points for FedAvg, 13.0 for the local forecasters, 15.4 for pooling and 16.1 at ε = 1. The cost ratio changed the levels: at 4:1, where the cost-optimal threshold is low enough that most at-risk lines were still flagged after the base rate rose, the FedAvg saving was 35% in the base scenario and 37% under the shock, whereas at 16:1 it fell from 25% to 14%; at ε = 1 the saving under the shock was 8% at 8:1 and -3% at 16:1. Across the seven logistic regimes of Table 8 other than client-level DP, the shock cost between 1.5 and 1.7 AUC points but between 8.5 and 16.1 points of relative saving at 8:1: calibration, not ranking, is what a shock breaks.

Table 13: Base and shock scenarios at τ = 0.5, K = 10, T = 10: pooled AUC, Brier score and relative cost saving at three cost ratios, by model and regime, including local recalibration

Model and regime	AUC base / shock	Brier base / shock	Saving 8:1 base / shock	Saving 4:1 base / shock	Saving 16:1 base / shock
LR: Local only	0.844 / 0.827	0.122 / 0.161	26% / 13%	26% / 28%	21% / 3%
LR: Centralised pooling	0.879 / 0.864	0.110 / 0.150	34% / 18%	35% / 33%	27% / 9%
LR: FedAvg	0.876 / 0.860	0.109 / 0.148	33% / 24%	35% / 37%	25% / 14%
LR: Sample-level DP, ε = 1	0.834 / 0.817	0.132 / 0.170	24% / 8%	24% / 24%	19% / -3%
LR: Sample-level DP, ε = 2	0.853 / 0.836	0.121 / 0.160	30% / 15%	30% / 29%	24% / 6%
LR: Sample-level DP, ε = 4	0.859 / 0.843	0.118 / 0.158	31% / 17%	32% / 31%	25% / 7%
LR: Client-level DP, ε = 16	0.747 / 0.727	0.232 / 0.274	-11% / -51%	-4% / -3%	-44% / -128%
LR: FedAvg + local recalibration	0.867 / 0.854	0.112 / 0.153	34% / 17%	35% / 32%	26% / 4%
LR: Sample-level DP ε = 2 + recalibration	0.842 / 0.830	0.124 / 0.164	28% / 9%	26% / 24%	22% / 0%
MLP: Local only	0.810 / 0.797	0.139 / 0.180	1% / -15%	9% / 11%	-11% / -39%
MLP: Centralised pooling	0.859 / 0.844	0.120 / 0.160	25% / 6%	28% / 25%	17% / -9%
MLP: FedAvg	0.845 / 0.832	0.126 / 0.166	24% / 10%	24% / 25%	18% / -1%
MLP: Sample-level DP, ε = 1	0.758 / 0.745	0.182 / 0.257	-68% / -137%	-32% / -56%	-113% / -224%
MLP: Sample-level DP, ε = 2	0.762 / 0.750	0.189 / 0.258	-79% / -136%	-44% / -60%	-115% / -209%
MLP: Sample-level DP, ε = 4	0.760 / 0.749	0.194 / 0.260	-92% / -140%	-52% / -62%	-128% / -210%
MLP: Client-level DP, ε = 16	0.740 / 0.729	0.244 / 0.294	-93% / -124%	-25% / -25%	-248% / -319%
MLP: FedAvg + local recalibration	0.837 / 0.824	0.131 / 0.171	18% / 2%	18% / 20%	12% / -14%
MLP: Sample-level DP ε = 2 + recalibration	0.760 / 0.747	0.161 / 0.205	-7% / -24%	-14% / -6%	-4% / -28%

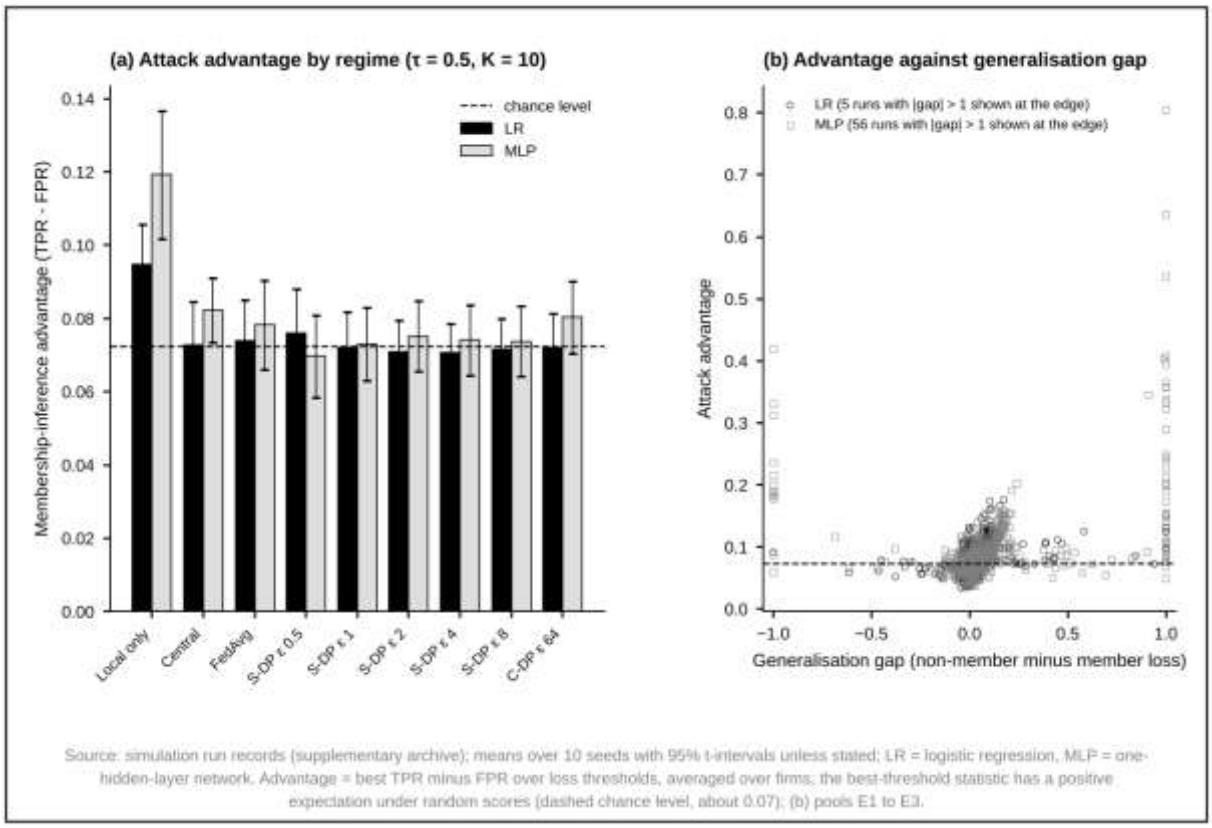
Empirical privacy against membership inference (RQ4)

Table 14 and Figure 8 report the loss-threshold attack. Because the attack statistic is the best advantage over all thresholds, it has a positive expectation even when the model leaks nothing; for firms of the simulated sizes, random scores give an expected advantage of 0.072 (standard deviation 0.014 across federations), which is the chance level against which the values must be read. Against that level, the shared logistic forecasters leaked nothing measurable: the FedAvg logistic forecaster had an advantage of 0.074 (95 per cent interval 0.063 to 0.085) and an attack AUC of 0.498, the centralised forecaster 0.073, and the sample-level private forecasters between 0.071 and 0.076. The local forecasters did leak: the local logistic forecasters had an advantage of 0.095 and the local networks 0.119 (attack AUC 0.533), with generalisation gaps of 0.067 and 0.128 against -0.002 and 0.004 for the FedAvg forecasters; the centralised network, trained on all panels without noise, leaked slightly (0.082). Across all 1,500 runs of E1 to E3 the attack advantage rose with the generalisation gap (Spearman ρ = 0.53 for logistic regression and 0.52 for the network), the relation the attack literature predicts (Yeom et al., 2018; Shokri et al., 2017); the largest advantage for the logistic forecasters was 0.176, in a local run, and the largest for the networks (0.804) occurred in a client-level DP (ε = 4) run with K = 5, whose model had diverged under the noise (pooled AUC 0.487, generalisation gap 41.0 nats) and for which the attack statistic is not meaningful.

Table 14: Loss-threshold membership inference against the released forecasters: attack advantage, attack AUC and generalisation gap by model and regime (τ = 0.5, K = 10); chance level of the advantage is about 0.07

Model and regime	Attack advantage (95% CI)	Attack AUC	Generalisation gap	Utility AUC
LR: Local only	0.095 (0.084 to 0.106)	0.517	0.0668	0.844
LR: Centralised pooling	0.073 (0.061 to 0.085)	0.498	0.0026	0.879
LR: FedAvg	0.074 (0.063 to 0.085)	0.498	-0.0023	0.876
LR: Sample-level DP, ε = 0.5	0.076 (0.064 to 0.088)	0.499	-0.0007	0.786
LR: Sample-level DP, ε = 1	0.072 (0.062 to 0.082)	0.498	-0.0027	0.834
LR: Sample-level DP, ε = 4	0.071 (0.063 to 0.079)	0.498	-0.0043	0.859
LR: Sample-level DP, ε = 8	0.072 (0.063 to 0.080)	0.498	-0.0046	0.861
LR: Client-level DP, ε = 64	0.072 (0.063 to 0.081)	0.498	0.0005	0.820
MLP: Local only	0.119 (0.102 to 0.137)	0.533	0.1284	0.810
MLP: Centralised pooling	0.082 (0.074 to 0.091)	0.506	0.0523	0.859
MLP: FedAvg	0.078 (0.066 to 0.090)	0.499	0.0040	0.845
MLP: Sample-level DP, ε = 0.5	0.070 (0.058 to 0.081)	0.495	-0.0086	0.728
MLP: Sample-level DP, ε = 1	0.073 (0.063 to 0.083)	0.498	0.0018	0.758
MLP: Sample-level DP, ε = 4	0.074 (0.064 to 0.084)	0.499	0.0017	0.760
MLP: Sample-level DP, ε = 8	0.074 (0.064 to 0.083)	0.499	0.0012	0.757
MLP: Client-level DP, ε = 64	0.080 (0.070 to 0.090)	0.502	0.0601	0.758

Figure 8: Membership-inference advantage by regime for both models, and advantage against the generalisation gap across all runs of E1 to E3



Two readings follow. First, for the low-capacity forecasters and short panels of this setting, the federated forecaster itself overfits little, and the simple attack cannot distinguish members from non-members whether or not differential privacy is applied; the formal budget buys protection against adversaries and attacks stronger than the one tested, not a measurable reduction of this attack's success, consistent with evaluations that found a wide gap between the formal ε and the empirical advantage of black-box attacks on well-generalising models (Jayaraman & Evans, 2019) and with audits showing that weak attacks leave the formal bound far from tight (Jagielski et al., 2020). Second, the regime that leaks most is the one with no sharing at all: a local forecaster trained on a hundred correlated records memorises them, and a firm that releases such a model to a platform, a partner or a software vendor exposes its order lines more than a participant in a federation does. The attack considered here is the weakest of its family; white-box attacks on the exchanged updates are stronger and are the threats that motivate sample-level DP and secure aggregation together (Nasr et al., 2019; Melis et al., 2019; Geiping et al., 2020).

Sensitivity and local recalibration

Table 15 and Figure 9 report the sensitivity analysis for the sample-level mechanism at ε = 2 (logistic regression, five seeds). The clipping norm and the number of local epochs interact: with a small clipping norm (C = 0.5) more local epochs helped in AUC (0.838 with one epoch, 0.848 with five), although the cost saving of that cell fell (28% to 10%), because each step moves less and the noise, proportional to C, is small; with a large clipping norm (C = 4) more epochs hurt (0.809 to 0.802), because the noise grows with C and the budget is spent over more steps. The best configurations of the grid in AUC were C = 1 with 2 epochs (0.849) and C = 0.5 with 5 epochs (0.848); the default used in the main experiments (C = 1, one epoch, 0.842) is close to the best, and the main results do not rest on an unfavourable configuration. The number of rounds mattered little under a fixed budget: 30, 60 and 120 rounds gave 0.845, 0.842 and 0.850, differences within the seed-to-seed variation, while FedAvg without noise moved from 0.864 to 0.874; the accountant charges every additional step against the same ε, so doubling the communication did not buy a corresponding gain (Abadi et al., 2016; McMahan et al., 2018).

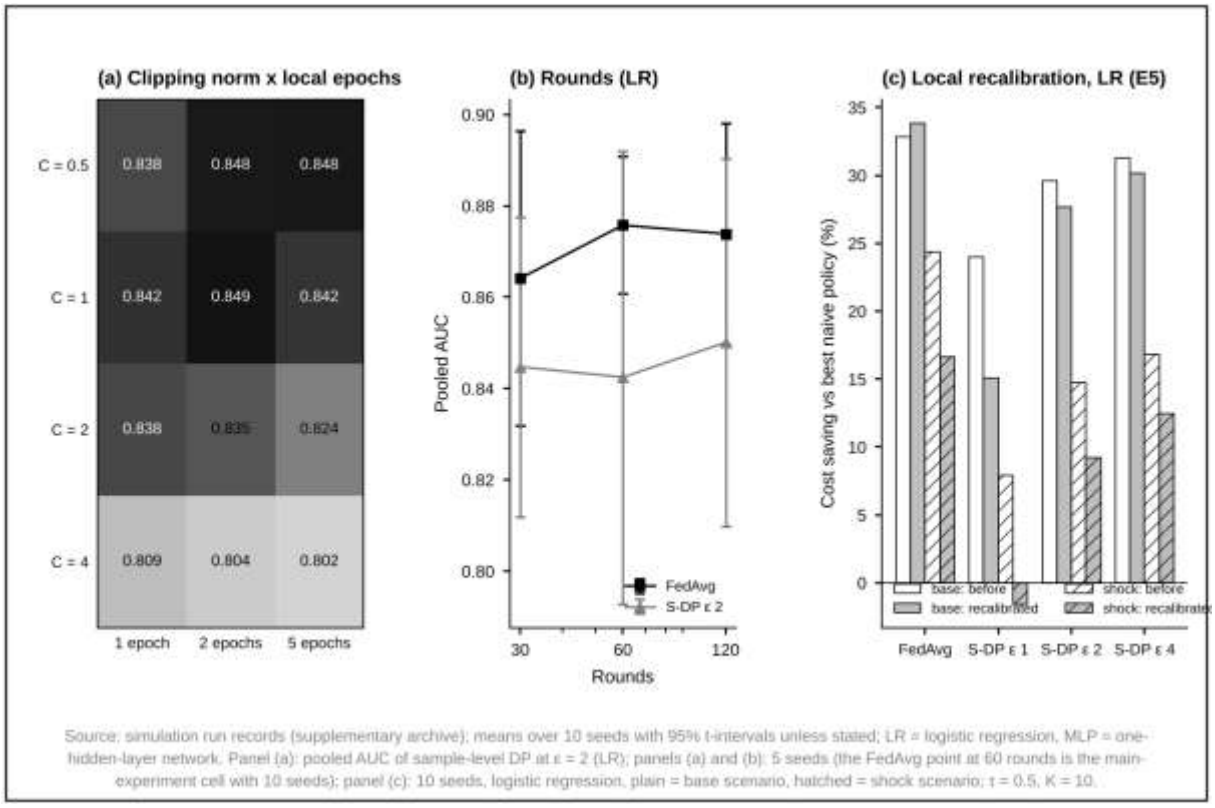
Local recalibration (E5) did not help in this setting, in contrast to its effect in the companion study of cross-sectional supplier risk. Fitting a monotone two-parameter rescaling of the shared forecaster's scores on each firm's own training records left the cost saving of the FedAvg logistic forecaster unchanged in the base scenario (33% to 34%) and lowered the point estimates for the private forecasters: at ε = 1 from 24% to 15%, at ε = 2 from 30% to 28% and at ε = 4 from 31% to 30%, none of these paired reductions being significant ($p = 0.333$ at ε = 1); the pooled AUC fell by about a point because the firms' scores were rescaled relative to one another (per-firm AUC is unchanged by construction), and under the shock the recalibrated forecasters lost more than the originals (17% against 24% for FedAvg; 9% against 15% at ε = 2). The explanation is the panel structure: a firm's own training records are few, correlated within lines and drawn from periods whose base rate differs from the forecast period, so that a rescaling fitted to them is itself miscalibrated for the period ahead, and when the regional factor moves it moves in the wrong direction. Where the companion study found calibration to be the part of the privacy cost that each participant could repair alone, here the repair needs information about the coming period that no single firm's history contains; monitoring the regional indicators and re-thresholding on recent realised disruptions, rather than refitting on the training panel, is the remedy the results point to (Elkan, 2001; Guo et al., 2017). For the network, whose private versions had negative savings, recalibration raised the saving at ε = 4 from -92% to -14% and at ε = 1 from -68% to -3%, still far below the logistic forecaster; the fallback to an intercept-only rescaling was needed in 0.1 of ten firms per federation on average at ε = 1 for the network and in 0.0 for logistic regression.

Table 15: Sensitivity of the sample-level mechanism to clipping norm, local epochs and rounds (ε = 2, logistic regression, five seeds; the FedAvg reference at 60 rounds is the main-experiment cell with ten seeds), and the effect of local recalibration (E5, ten seeds, base scenario)

Configuration	AUC pooled (95% CI)	Cost saving	Brier
Sample-level DP ε = 2: clipping C = 0.5, local epochs 1	0.838 (0.798 to 0.878)	28%	0.144
Sample-level DP ε = 2: clipping C = 0.5, local epochs 2	0.848 (0.798 to 0.898)	28%	0.136
Sample-level DP ε = 2: clipping C = 0.5, local epochs 5	0.848 (0.794 to 0.902)	10%	0.137
Sample-level DP ε = 2: clipping C = 1, local epochs 1	0.842 (0.793 to 0.892)	28%	0.133
Sample-level DP ε = 2: clipping C = 1, local epochs 2	0.849 (0.800 to 0.899)	25%	0.131
Sample-level DP ε = 2: clipping C = 1, local epochs 5	0.842 (0.779 to 0.905)	-11%	0.148
Sample-level DP ε = 2: clipping C = 2, local epochs 1	0.838 (0.779 to 0.898)	25%	0.134
Sample-level DP ε = 2: clipping C = 2, local epochs 2	0.835 (0.781 to 0.889)	13%	0.142
Sample-level DP ε = 2: clipping C = 2, local epochs 5	0.824 (0.748 to 0.900)	-44%	0.166
Sample-level DP ε = 2: clipping C = 4, local epochs 1	0.809 (0.739 to 0.879)	10%	0.151
Sample-level DP ε = 2: clipping C = 4, local epochs 2	0.804 (0.729 to 0.880)	-24%	0.168
Sample-level DP ε = 2: clipping C = 4, local epochs 5	0.802 (0.715 to 0.888)	-79%	0.192
Sample-level DP ε = 2: 30 rounds (C = 1, 1 epoch)	0.845 (0.812 to 0.878)	27%	0.136
FedAvg: 30 rounds	0.864 (0.832 to 0.896)	27%	0.124
Sample-level DP ε = 2: 60 rounds (C = 1, 1 epoch)	0.842 (0.793 to 0.892)	28%	0.133
FedAvg: 60 rounds	0.876 (0.861 to 0.891)	33%	0.109
Sample-level DP ε = 2: 120 rounds (C = 1, 1 epoch)	0.850 (0.810 to 0.890)	25%	0.129
FedAvg: 120 rounds	0.874 (0.850 to 0.898)	34%	0.120
LR: FedAvg	0.876 (0.861 to 0.891)	33%	0.109
LR: FedAvg + local recalibration	0.867 (0.846 to 0.888)	34%	0.112
LR: Sample-level DP, ε = 1	0.834 (0.804 to 0.864)	24%	0.132
LR: Sample-level DP ε = 1 + recalibration	0.824 (0.782 to 0.866)	15%	0.133
LR: Sample-level DP, ε = 2	0.853 (0.828 to 0.878)	30%	0.121
LR: Sample-level DP ε = 2 + recalibration	0.842 (0.806 to 0.878)	28%	0.124
LR: Sample-level DP, ε = 4	0.859 (0.835 to 0.883)	31%	0.118
LR: Sample-level DP ε = 4 + recalibration	0.848 (0.816 to 0.881)	30%	0.121
MLP: FedAvg	0.845 (0.821 to 0.869)	24%	0.126
MLP: FedAvg + local recalibration	0.837 (0.808 to 0.867)	18%	0.131
MLP: Sample-level DP, ε = 1	0.758 (0.707 to 0.809)	-68%	0.182
MLP: Sample-level DP ε = 1 + recalibration	0.755 (0.701 to 0.809)	-3%	0.159
MLP: Sample-level DP, ε = 2	0.762 (0.710 to 0.814)	-79%	0.189
MLP: Sample-level DP ε = 2 + recalibration	0.760 (0.704 to 0.815)	-7%	0.161

Configuration	AUC pooled (95% CI)	Cost saving	Brier
MLP: Sample-level DP, ε = 4	0.760 (0.703 to 0.817)	-92%	0.194
MLP: Sample-level DP ε = 4 + recalibration	0.758 (0.698 to 0.817)	-14%	0.163

Figure 9: Sensitivity of the sample-level mechanism to clipping norm and local epochs, to the number of rounds, and the effect of local recalibration on decision value in the base and shock scenarios



Source: simulation run records (supplementary archive); means over 10 seeds with 95% t-intervals unless stated; LR = logistic regression, MLP = one-hidden-layer network. Panel (a): pooled AUC of sample-level DP at ε = 2 (LR); panels (a) and (b): 5 seeds (the FedAvg point at 60 rounds is the main-experiment cell with 10 seeds); panel (c): 10 seeds, logistic regression, plain = base scenario, hatched = shock scenario; τ = 0.5, K = 10.

DESIGN GUIDANCE FOR A PRIVACY-PRESERVING MATERIALS FEDERATION

From trade-off curves to design decisions

The results reduce to a small number of design decisions that a group of construction firms faces when it sets up a shared disruption forecaster, and Table 16 maps each decision to what the simulation showed and to a prescription. The first decision is the objective: because privacy noise and regional shocks both degrade calibration faster than ranking, options should be compared on the expected cost of the buffering decision at the consortium's cost ratio, in calm and in shock periods, and on watch-list capture, not on AUC alone. The second is scope: federation is worth most to small firms and to firms with short histories, and when members differ strongly in material mix and exposure, the shared forecaster should be fine-tuned locally, which improved it at moderate and strong heterogeneity and cost a tenth of a point with identical firms. The third and fourth concern the two guarantees: sample-level differential privacy at ε between 2 and 8 is affordable for a low-dimensional forecaster and should be the default protection of individual order-line records, with the unit of protection (a line-period) stated plainly, whereas a client-level guarantee at a conventional budget (ε = 4) left the forecaster below local training at every federation size tested, including 100 firms (0.823 against 0.848), and the client-level forecaster came within a point of FedAvg only at ε = 16 with 100 firms (0.9 points) or at ε = 64 with 50 or more firms, budgets that offer little formal protection, so that for regional consortia they must be replaced by secure aggregation, contractual controls and audits of the released model. The fifth is the model: with about a hundred correlated records per firm, logistic regression matched or beat the network on every utility measure, with and without noise, except under client-level DP at ε = 4, where

neither was usable. The sixth is calibration under shocks: local recalibration on the firm's own panel did not repair the private forecasters and worsened them under a shock, so that thresholds must be maintained against recent realised disruptions and regional indicators rather than refitted on history. The seventh is history: the value of federation and the cost of privacy both fall as histories lengthen, which argues for recording line-level delivery performance before the federation is needed. The eighth is audit: the attack test is a check on the released model, not a substitute for the guarantee.

Table 16: Design decisions for a cross-firm materials-disruption federation, the simulation evidence for each and the prescription, with supporting literature

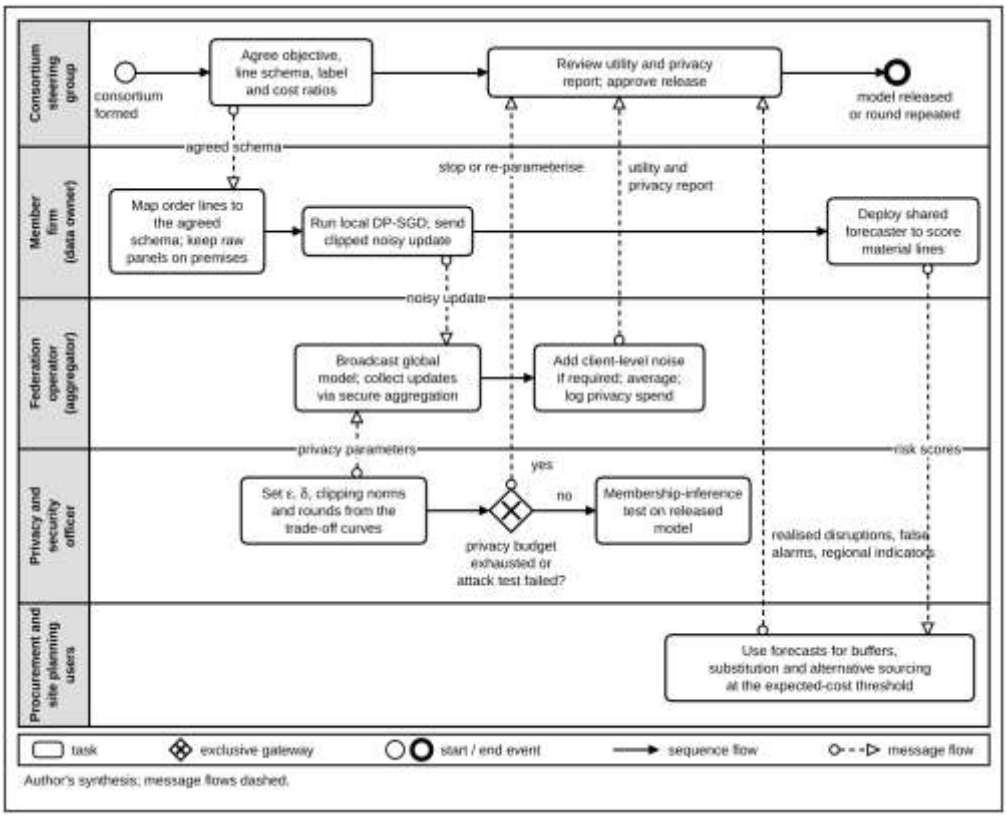
Design decision	What the simulation shows	Prescription	Supporting literature
1 Objective and cost ratio	Cost saving against the best naive policy is 33% for FedAvg and 24% at ε = 1 (LR, 8:1); it is 35% at 4:1 and 25% at 16:1; decision value falls faster than AUC	Fix the cost ratio and the watch-list size first; evaluate every option on expected cost at that ratio, not on AUC alone	(Elkan, 2001; Hand, 2009)
2 Federation scope	Federation gains 3.2 AUC points over local models at τ = 0.5 (LR), most for small firms; at τ = 1 local models reach a mean per-firm AUC of 0.849 against 0.845 for FedAvg and 0.865 after fine-tuning	Form federations of firms in one region with related material mixes; use local fine-tuning when members differ strongly	(Li et al., 2020c; Smith et al., 2017; Fallah et al., 2020)
3 Protection of records	Sample-level DP at ε = 4 costs 1.7 points and at ε = 1 costs 4.2 points (LR, T = 10)	Use DP-SGD at each firm with ε of 2 to 8 and δ = 1e-5 as the default record-level guarantee; calibrate σ per firm; state that the unit is a line-period	(Abadi et al., 2016; McMahan et al., 2018; Mironov, 2017)
4 Protection of firms	Client-level DP at ε = 16 costs 12.9 points with K = 10 and 0.9 with K = 100 (LR)	Treat client-level guarantees as attainable only in large federations; otherwise protect firm aggregates by secure aggregation, contract and audit	(McMahan et al., 2018; Wei et al., 2020; Bonawitz et al., 2017)
5 Model choice	Under sample-level DP at ε = 4 the MLP loses 8.6 points against 1.7 for LR	Prefer low-dimensional interpretable forecasters under privacy noise and short panels; add capacity only when histories justify it	(Bassily et al., 2014; Chaudhuri et al., 2011; Lessmann et al., 2015)
6 Calibration and shocks	A regional shock at the horizon cuts the saving of the FedAvg LR model from 33% to 24% and of local models from 26% to 13%; local recalibration on each firm's own panel lowers the saving at ε = 1 from 24% to 15% and worsens the shock penalty	Do not refit thresholds on the training panel; monitor the regional indicators and recent realised disruptions and re-threshold when the base rate moves	(Guo et al., 2017; Elkan, 2001; Alsharef et al., 2021)
7 History length	With T = 6 periods the loss at ε = 1 is 5.7 points and local models are 5.0 points below FedAvg; with T = 24 the figures are 1.4 and 0.4 (LR)	Start recording line-level delivery performance before the federation is needed; short histories make federation most valuable and privacy most expensive	(Bagdasaryan et al., 2019; Wang et al., 2019; Thunberg & Persson, 2014)

Design decision	What the simulation shows	Prescription	Supporting literature
8 Empirical audit	Membership-inference advantage is 0.074 for FedAvg and 0.072 at ε = 1 (LR); 0.119 for local MLPs	Run attack tests on every released model; use them to justify budgets but not to replace the formal guarantee	(Yeom et al., 2018; Jayaraman & Evans, 2019; Jagielski et al., 2020)

Governance workflow

Figure 10 lays out the decisions as a workflow with five roles. A steering group of the member firms agrees the objective, the common line schema (material category, supplier, order and delivery fields), the disruption label and the cost ratios, and later reviews a utility-and-privacy report before a forecaster is released. Each member maps its order lines to the schema and keeps the raw panel on its own systems, runs local DP-SGD and sends clipped, noised updates, and later deploys the released forecaster to score its lines. A federation operator broadcasts the model, collects updates through secure aggregation, adds client-level noise only if the federation is large enough to afford it, averages, and keeps a ledger of the privacy budget spent (Dwork et al., 2010; Kairouz et al., 2017; Mironov, 2017). A privacy and security officer sets ε, δ, the clipping norms and the number of rounds from trade-off curves measured on the consortium's own panels, runs membership-inference tests on each released forecaster, and stops or re-parameterises the process when the budget is exhausted or a test fails. Procurement and site-planning users apply the forecasts to buffers, substitution and alternative sourcing at the expected-cost threshold and return realised disruptions, false alarms and the regional indicators that drive re-thresholding. The arrangement gives the information-sharing questions of the construction literature, who sees what and who decides, technical and organisational answers at the same time (Akintoye et al., 2000; Xue et al., 2007; Klein & Rai, 2009; Fawcett et al., 2008).

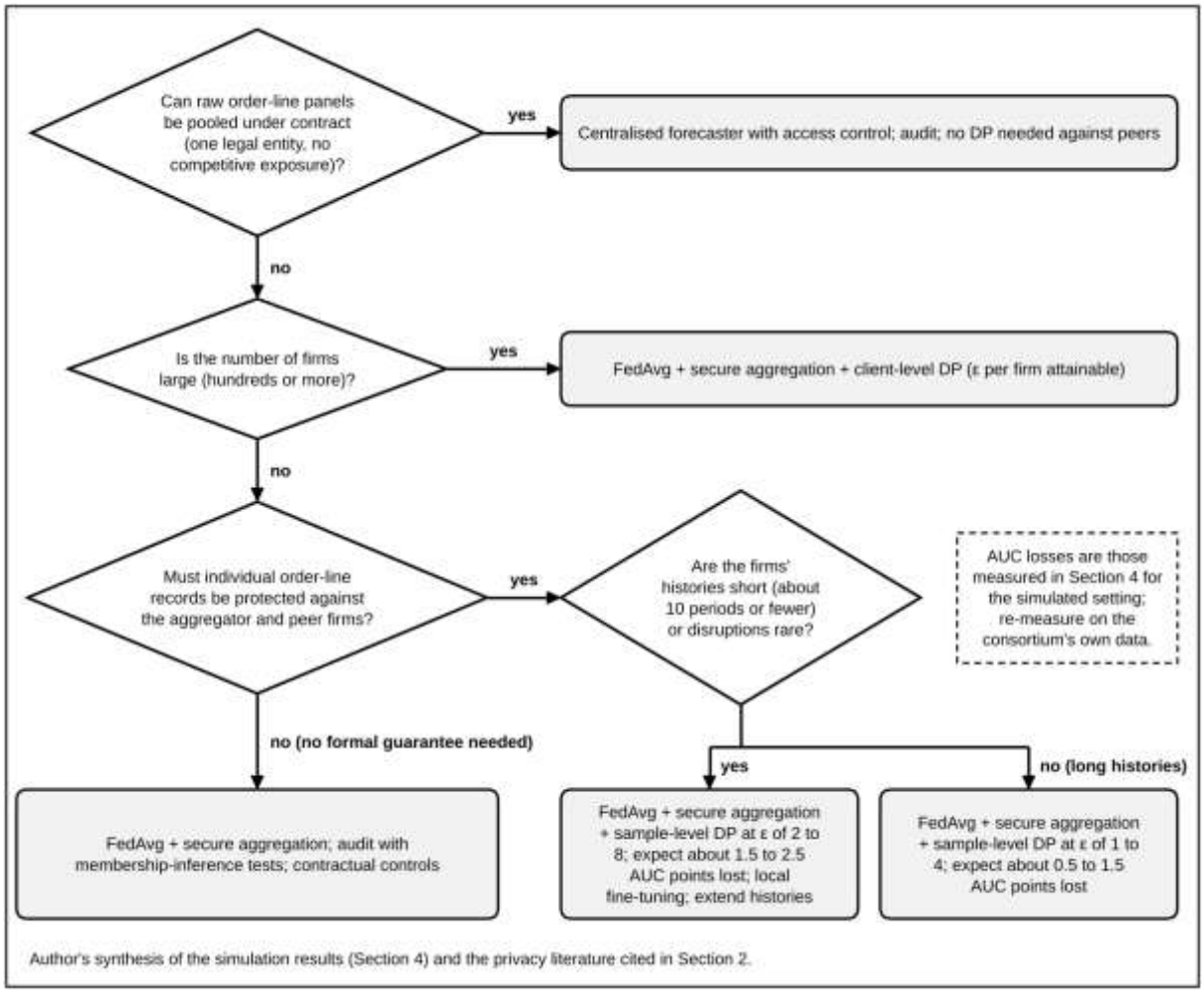
Figure 10: Governance workflow for a federated materials-disruption consortium, with five roles and the message flows between them



Choosing the mechanism

Figure 11 states the choice of mechanism as a decision tree. When order-line panels can be pooled under one legal entity with no competitive exposure, for example within a group of companies, a centralised forecaster with access control and audit is simplest and best. When they cannot, the number of firms decides whether client-level guarantees are attainable: with hundreds of members, as on a national procurement platform, FedAvg with secure aggregation and client-level DP is the arrangement the literature has demonstrated at scale (McMahan et al., 2018; Bonawitz et al., 2019); with tens of members it is not. When individual order-line records must be protected formally against the aggregator and peers, sample-level DP-SGD with secure aggregation is the mechanism, with the budget chosen by the length of history: for members with short histories, around ε of 2 to 8 with an expected loss of a few AUC points, local fine-tuning, and a plan to extend the panels; for members with long histories, around ε of 1 to 4 with a smaller loss. When no formal record-level guarantee is required, FedAvg with secure aggregation, audited with attack tests and governed by contract, is the pragmatic choice, with the reservation that the released forecaster is then protected only empirically. The AUC losses in the tree are those measured in this simulation and must be re-measured on a consortium's own panels.

Figure 11: Decision tree for choosing the privacy mechanism of a cross-firm materials-disruption forecaster

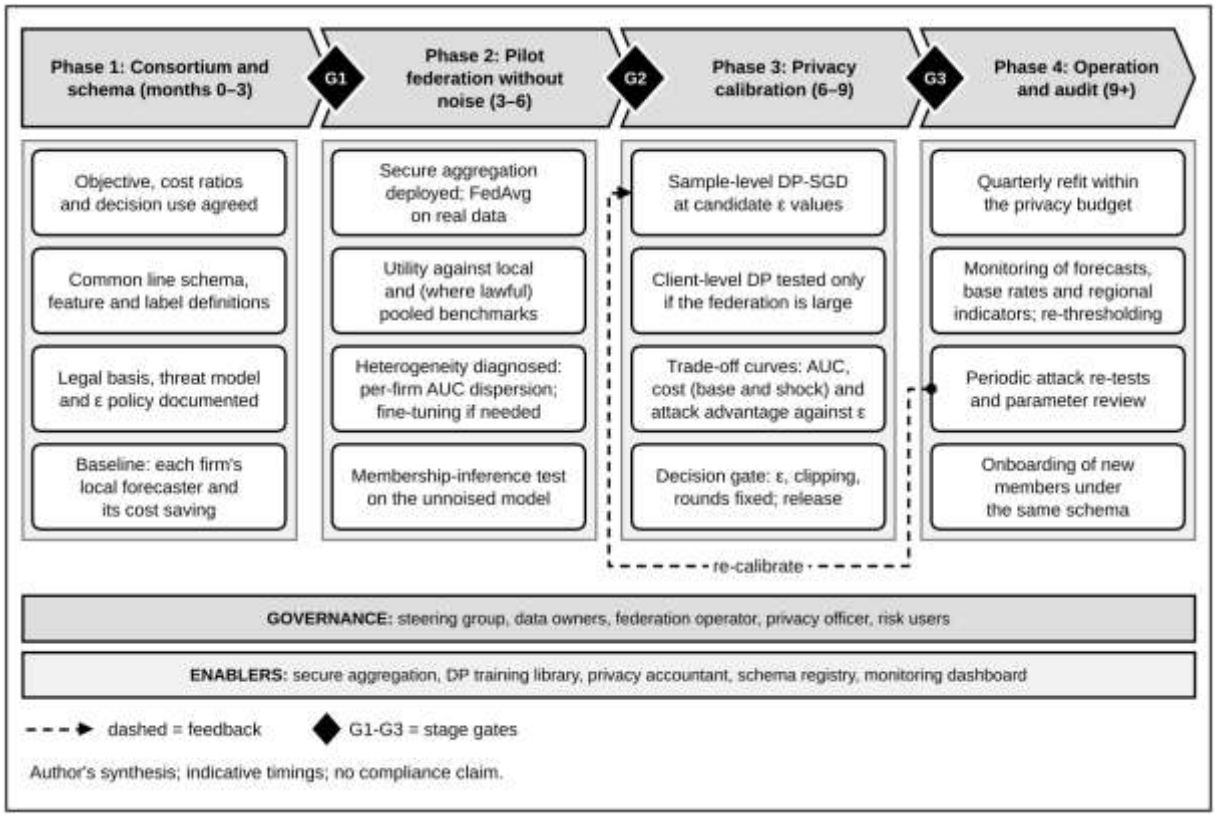


Implementation roadmap

Figure 12 arranges the steps in four phases with stage gates. The first phase forms the consortium and the schema: objective, cost ratios, line and label definitions, legal basis, threat model and a written ε policy, and a baseline consisting of each member's local forecaster and its cost saving, which is the status quo against which the federation must prove itself. The second phase pilots the federation without noise, with secure aggregation deployed, measures utility against local and, where lawful,

pooled benchmarks, diagnoses heterogeneity from the dispersion of per-member performance and adds fine-tuning where it is large, and runs the first attack tests. The third phase calibrates privacy: sample-level DP at candidate budgets, client-level DP only if the federation is large, trade-off curves for AUC, cost in calm and shock periods and attack advantage against ε, and a decision gate at which the budget, the clipping norms and the number of rounds are fixed and the forecaster is released. The fourth phase operates the federation: quarterly refits within the privacy budget, monitoring of forecasts, base rates and regional indicators with re-thresholding when the base rate moves, periodic attack re-tests and parameter reviews, and the onboarding of new members under the same schema. The timings are indicative; the simulation says nothing about them.

Figure 12: Implementation roadmap for a privacy-preserving materials-disruption federation, in four phases with stage gates



DISCUSSION

Principal findings

The study set out to quantify what a group of construction firms gains and loses when it trains a shared materials-disruption forecaster with federated learning and differential privacy. Four findings answer the research questions. First, federation recovers almost the whole benefit of pooling when firms are similar (0.3 points from the centralised benchmark at $\tau = 0.5$ for logistic regression), gives small firms and firms with short histories forecasters they could not estimate alone (the worst-served firm improved from an AUC of 0.780 to 0.837; with six periods of history local forecasters were 5.0 points below FedAvg), and when firms differ strongly, local fine-tuning of the shared forecaster beats both the shared and the local models. Second, sample-level differential privacy is affordable for a low-dimensional forecaster but costs more than with larger data sets: $\varepsilon = 4$ cost 1.7 AUC points and $\varepsilon = 1$ cost 4.2 points with ten periods of history, whereas client-level differential privacy at conventional budgets destroyed the forecaster with ten firms (19.6 points at $\varepsilon = 4$) and still left it 5.9 points below FedAvg and 2.6 below local training with a hundred. Third, a regional shock at the forecast horizon barely moved discrimination but cut decision value through calibration (the FedAvg saving fell from 33% to 24%, the local saving from 26% to 13%), the federated forecaster kept more of its value under the shock than isolated local ones, and local recalibration on the firm's own panel, which repaired the

private models in the companion study, did not help here and worsened the shock penalty. Fourth, the simple membership-inference attack learned nothing from the shared logistic forecasters with or without differential privacy and something from isolated local ones (advantage 0.095 against a chance level of 0.072; 0.119 for local networks).

These findings are consistent with, and sharpen, results from other domains. The dependence of client-level DP on the number of participants is the central message of the original DP-FedAvg work and of the convergence analysis of DP federated learning (McMahan et al., 2018; Wei et al., 2020); the simulation shows what it means for regional consortia. The advantage of low-dimensional models under privacy noise follows from the dimension dependence of private empirical risk minimisation (Bassily et al., 2014; Chaudhuri et al., 2011) and from empirical evaluations (Jayaraman & Evans, 2019). The gap between formal ε and the success of black-box attacks has been reported for well-generalising models (Jayaraman & Evans, 2019; Jagielski et al., 2020). The interaction of heterogeneity with shared models is the subject of the personalisation literature (Li et al., 2020c; Smith et al., 2017; Fallah et al., 2020). What the study adds is the joint quantification in a setting the construction supply-chain literature needs and has not measured, with a shared regional shock process, panel data and decision-relevant metrics, and the finding that calibration under shocks is where a shared forecaster earns its keep and where local repair fails.

Implications for research

Three directions follow. The first is realism: the simulation should be followed by a federation on real order-line panels held by a small number of cooperating contractors or merchants, where the generating process is unknown, the features are what firms actually record and the realised disruptions are the labels; such a study could use the same regimes, metrics and attack tests and would establish whether the trade-off curves transfer (Brintrup et al., 2020; Gondia et al., 2020). The second is method: the unit of the sample-level guarantee here is a line-period, and a guarantee at the level of a line or a supplier relationship requires group privacy or user-level accounting, whose utility cost for short panels should be measured; sample-level DP with local fine-tuning deserves a formal treatment of its privacy properties; and calibration under distribution shift, which the results identify as the weak point, invites methods that update thresholds from recent outcomes or from regional indicators without refitting on confidential history (Li et al., 2020c; Karimireddy et al., 2020; Hu et al., 2020; Guo et al., 2017). Vertical federations, in which contractors, merchants and manufacturers hold different features of the same lines, are a natural extension (Cheng et al., 2021). The third is measurement: the attack used here is the weakest of its family, and stronger white-box and auditing attacks should be run against the exchanged updates as well as the final forecaster to estimate empirical lower bounds on ε in this setting (Nasr et al., 2019; Jagielski et al., 2020; Song & Mittal, 2021); the economic framing of ε as a choice between the cost of disclosure and the cost of worse decisions could be made concrete with the cost ratios of construction procurement (Hsu et al., 2014; Ghosh et al., 2012).

Implications for practice

For a group of firms considering a shared disruption forecaster, the results suggest a sequence. Agree the decision and its costs first, because the forecaster will be judged on expected cost in calm and in shock periods, not on AUC. Start with firms in one region whose material mixes are related, deploy secure aggregation and federated averaging without noise, measure the gain over each member's local forecaster, and add local fine-tuning. Protect individual records with sample-level differential privacy at a budget chosen from the members' history lengths, prefer logistic regression or similarly low-dimensional forecasters, and state that the unit of protection is a line-period. Do not refit thresholds on the training panel; maintain them against recent realised disruptions and regional indicators, and expect decision value to fall in a shock even when ranking holds. Do not promise members a formal client-level guarantee at a conventional budget unless the federation has many tens of members; rely instead on secure aggregation, contracts and attack audits. Treat the local status quo as the baseline for privacy as well as utility: an isolated forecaster trained on a hundred correlated records and shared with a platform or a vendor leaks more than a federated one.

Limitations

The study has the limitations of a simulation. The data are synthetic: the features, their correlations, the risk relationships, the latent processes, the heterogeneity structure and the shock were set by design to

realism targets taken from the literature, not estimated from confidential panels, and the absolute levels of AUC, cost saving and attack advantage describe the simulated world; the comparisons between regimes are the findings, and even these may change with real feature sets, label definitions and dependence structures the generator does not contain, such as suppliers shared by several firms, several regions, or shocks that change the risk relationships rather than the base rate. The shock scenario imposes one fixed adverse regional path irrespective of the realised one, which in four of ten federations of the main cell was already above the imposed origin value, so that the averaged shock penalties understate what a shock does to a federation in calm conditions (Section 4.5 reports that subset). The models are small by the standards of the federated learning literature, which is appropriate for the data volumes considered but means that results for larger models and longer histories were not measured beyond 24 periods. The sample-level guarantee protects a line-period, not a line; a line-level guarantee would require group privacy and a larger budget. The privacy accounting uses the standard Rényi conversion, and the client-level mechanism without participant sampling forgoes amplification that a large federation could use. The attack is a single black-box attack on the final forecaster. Communication, system heterogeneity, dropouts, malicious participants and incentives were outside the scope. Ten seeds per cell give confidence intervals of about one AUC point for the main comparisons and wider ones for the cost savings, which depend on a threshold; several of the differences reported between adjacent budgets are within that uncertainty, and the many comparisons are descriptive. All references are journal and conference publications with a version of record of 2021 or earlier, as required by the author's constraint, so that later work on federated learning for industrial data and on construction supply chains after the pandemic is not cited.

CONCLUSION

Construction firms that would benefit from a shared forecaster of materials disruption cannot pool the order histories that would make it good, and federated learning with differential privacy is the arrangement that promises to learn across them without sharing. This study measured what that arrangement costs in a controlled setting that resembles a regional consortium of contractors and merchants rather than a fleet of devices, with short correlated panels, rare disruptions and a regional shock process shared by every member. The answer is specific. Federated averaging recovers nearly all of the pooling benefit when members are similar, most helps the small members and those with short histories, and with local fine-tuning beats both the shared and the local forecasters when members differ. Record-level differential privacy at budgets between 2 and 8 costs a few points of AUC for a logistic forecaster on ten periods of history and less on longer ones; the same budgets cripple a neural network at these data volumes; and firm-level differential privacy at conventional budgets is not attainable with tens of participants, so that consortia of that size must protect their members' aggregates by secure aggregation, contract and audit rather than by noise. A regional shock at the forecast horizon leaves ranking almost intact and cuts decision value through calibration; the federated forecaster, which has learned the regional response from every member, keeps more of its value than isolated local ones, and local recalibration on a firm's own short history does not repair the loss. A simple membership-inference attack learns nothing from the shared logistic forecasters and something from isolated local ones. These results, the workflow, the decision tree and the roadmap derived from them are offered as a starting point for consortia and for the field study on real order-line panels that should follow.

REFERENCES

[1]. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS)* (pp. 308–318). ACM. https://doi.org/10.1145/2976749.2978318

[2]. Aditya, D., & Ahmed, H. C. H. R. (2022). Decision Support Systems for Last-Mile Logistics Optimization: A Data-Driven Framework for e-Commerce Fulfillment Networks. *International Journal of Scientific Interdisciplinary Research*, *12*(08), 01–49. https://doi.org/10.63125/h46vp562

[3]. Agarwal, N., Suresh, A. T., Yu, F. X., Kumar, S., & McMahan, B. (2018). cpSGD: Communication-efficient and differentially-private distributed SGD. In *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)* (pp. 7564–7575). Curran Associates. https://papers.nips.cc/paper/2018/hash/21ce689121e39821d07d04faab328370-Abstract.html

[4]. Akintoye, A., McIntosh, G., & Fitzgerald, E. (2000). A survey of supply chain collaboration and management in the UK construction industry. *European Journal of Purchasing & Supply Management*, *6*(3-4), 159–168. https://doi.org/10.1016/S0969-7012(00)00012-5

[5]. Aloini, D., Dulmin, R., Mininno, V., & Ponticelli, S. (2012). Supply chain management: A review of implementation risks in the construction industry. *Business Process Management Journal*, *18*(5), 735–761. https://doi.org/10.1108/14637151211270135

[6]. Alsharef, A., Banerjee, S., Uddin, S. M. J., Albert, A., & Jaselskis, E. J. (2021). Early impacts of the COVID-19 pandemic on the United States construction industry. *International Journal of Environmental Research and Public Health*, *18*(4), Article 1559. https://doi.org/10.3390/ijerph18041559

[7]. Anand, K. S., & Goyal, M. (2009). Strategic information management under leakage in a supply chain. *Management Science*, *55*(3), 438–452. https://doi.org/10.1287/mnsc.1080.0930

[8]. Ashuri, B., & Lu, J. (2010). Time series analysis of ENR construction cost index. *Journal of Construction Engineering and Management*, *136*(11), 1227–1237. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000231

[9]. Assaad, R., & El-adaway, I. H. (2021). Guidelines for responding to COVID-19 pandemic: Best practices, impacts, and future research directions. *Journal of Management in Engineering*, *37*(3), Article 06021001. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000906

[10]. Assaf, S. A., & Al-Hejji, S. (2006). Causes of delay in large construction projects. *International Journal of Project Management*, *24*(4), 349–357. https://doi.org/10.1016/j.ijproman.2005.11.010

[11]. Atallah, M. J., Elmongui, H. G., Deshpande, V., & Schwarz, L. B. (2003). Secure supply-chain protocols. In *IEEE International Conference on E-Commerce (CEC 2003)* (pp. 293–302). IEEE. https://doi.org/10.1109/COEC.2003.1210264

[12]. Ateniese, G., Mancini, L. V., Spognardi, A., Villani, A., Vitali, D., & Felici, G. (2015). Hacking smart machines with smarter ones: How to extract meaningful data from machine learning classifiers. *International Journal of Security and Networks*, *10*(3), 137–150. https://doi.org/10.1504/IJSN.2015.071829

[13]. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, *54*(6), 627–635. https://doi.org/10.1057/palgrave.jors.2601545

[14]. Bagdasaryan, E., Poursaeed, O., & Shmatikov, V. (2019). Differential privacy has disparate impact on model accuracy. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)* (pp. 15479–15488). Curran Associates. https://papers.nips.cc/paper/2019/hash/fc0de4e0396fff257ea362983c2dda5a-Abstract.html

[15]. Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., & Shmatikov, V. (2020). How to backdoor federated learning. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 108* (pp. 2938–2948). PMLR. https://proceedings.mlr.press/v108/bagdasaryan20a.html

[16]. Balle, B., & Wang, Y.-X. (2018). Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In *Proceedings of the 35th International Conference on Machine Learning (ICML), PMLR 80* (pp. 394–403). PMLR. https://proceedings.mlr.press/v80/balle18a.html

[17]. Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply chain risk management and artificial intelligence: State of the art and future research directions. *International Journal of Production Research*, *57*(7), 2179–2202. https://doi.org/10.1080/00207543.2018.1530476

[18]. Bassily, R., Smith, A., & Thakurta, A. (2014). Private empirical risk minimization: Efficient algorithms and tight error bounds. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS)* (pp. 464–473). IEEE. https://doi.org/10.1109/FOCS.2014.56

[19]. Behera, P., Mohanty, R. P., & Prakash, A. (2015). Understanding construction supply chain management. *Production Planning & Control*, *26*(16), 1332–1350. https://doi.org/10.1080/09537287.2015.1045953

[20]. Bhagoji, A. N., Chakraborty, S., Mittal, P., & Calo, S. (2019). Analyzing federated learning through an adversarial lens. In *Proceedings of the 36th International Conference on Machine Learning (ICML), PMLR 97* (pp. 634–643). PMLR. https://proceedings.mlr.press/v97/bhagoji19a.html

[21]. Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., & Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, *30*(3), 500–521. https://doi.org/10.1016/j.aei.2016.07.001

[22]. Blanchard, P., El Mhamdi, E. M., Guerraoui, R., & Stainer, J. (2017). Machine learning with adversaries: Byzantine tolerant gradient descent. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (pp. 119–129). Curran Associates. https://papers.nips.cc/paper/2017/hash/f4b9ec30ad9f68f89b29639786cb62ef-Abstract.html

[23]. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konečný, J., Mazzocchi, S., McMahan, H. B., Van Overveldt, T., Petrou, D., Ramage, D., & Roselander, J. (2019). Towards federated learning at scale: System design. In *Proceedings of Machine Learning and Systems 1 (SysML 2019)* (pp. 374–388). mlsys.org. https://proceedings.mlsys.org/paper_files/paper/2019/hash/7b770da633baf74895be22a8807f1a8f-Abstract.html

[24]. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS)* (pp. 1175–1191). ACM. https://doi.org/10.1145/3133956.3133982

[25]. Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

[26]. Brickell, J., & Shmatikov, V. (2008). The cost of privacy: Destruction of data-mining utility in anonymized data publishing. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 70–78). ACM. https://doi.org/10.1145/1401890.1401904

[27]. Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

[28]. Brintrup, A., Pak, J., Ratiney, D., Pearce, T., Wichmann, P., Woodall, P., & McFarlane, D. (2020). Supply chain data analytics for predicting supplier disruptions: A case study in complex asset manufacturing. *International Journal of Production Research*, *58*(11), 3330–3341. https://doi.org/10.1080/00207543.2019.1685705

[29]. Caldas, C. H., Menches, C. L., Reyes, P. M., Navarro, L., & Vargas, D. M. (2015). Materials management practices in the construction industry. *Practice Periodical on Structural Design and Construction*, *20*(3), Article 04014039. https://doi.org/10.1061/(ASCE)SC.1943-5576.0000238

[30]. Chaudhuri, K., Monteleoni, C., & Sarwate, A. D. (2011). Differentially private empirical risk minimization. *Journal of Machine Learning Research*, *12*, 1069–1109. https://jmlr.org/papers/v12/chaudhuri11a.html

[31]. Cheng, J. C. P., Law, K. H., Björnsson, H. C., Jones, A. T., & Sriram, R. D. (2010). A service oriented framework for construction supply chain integration. *Automation in Construction*, *19*(2), 245–260. https://doi.org/10.1016/j.autcon.2009.10.003

[32]. Cheng, K., Fan, T., Jin, Y., Liu, Y., Chen, T., Papadopoulos, D., & Yang, Q. (2021). SecureBoost: A lossless federated learning framework. *IEEE Intelligent Systems*, *36*(6), 87–98. https://doi.org/10.1109/MIS.2021.3082561

[33]. Clifton, C., Iyer, A., Cho, R., Jiang, W., Kantarcıoğlu, M., & Vaidya, J. (2008). An approach to securely identifying beneficial collaboration in decentralized logistics systems. *Manufacturing & Service Operations Management*, *10*(1), 108–125. https://doi.org/10.1287/msom.1070.0167

[34]. Darko, A., Chan, A. P. C., Adabre, M. A., Edwards, D. J., Hosseini, M. R., & Ameyaw, E. E. (2020). Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities. *Automation in Construction*, *112*, Article 103081. https://doi.org/10.1016/j.autcon.2020.103081

[35]. Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*(1), 1–30. https://www.jmlr.org/papers/v7/demsar06a.html

[36]. Dinur, I., & Nissim, K. (2003). Revealing information while preserving privacy. In *Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2003)* (pp. 202–210). ACM. https://doi.org/10.1145/773153.773173

[37]. Dubois, A., & Gadde, L.-E. (2002). The construction industry as a loosely coupled system: Implications for productivity and innovation. *Construction Management and Economics*, *20*(7), 621–631. https://doi.org/10.1080/01446190210163543

[38]. Duchi, J. C., Jordan, M. I., & Wainwright, M. J. (2013). Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science (FOCS)* (pp. 429–438). IEEE. https://doi.org/10.1109/FOCS.2013.53

[39]. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, *9*(3-4), 211–487. https://doi.org/10.1561/0400000042

[40]. Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography: TCC 2006, Lecture Notes in Computer Science 3876* (pp. 265–284). Springer. https://doi.org/10.1007/11681878_14

[41]. Dwork, C., Rothblum, G. N., & Vadhan, S. (2010). Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science (FOCS)* (pp. 51–60). IEEE. https://doi.org/10.1109/FOCS.2010.12

[42]. Dwork, C., Smith, A., Steinke, T., & Ullman, J. (2017). Exposed! A survey of attacks on private data. *Annual Review of Statistics and Its Application*, *4*(1), 61–84. https://doi.org/10.1146/annurev-statistics-060116-054123

[43]. Dwork, C. (2008). Differential privacy: A survey of results. In *Theory and Applications of Models of Computation: TAMC 2008, Lecture Notes in Computer Science 4978* (pp. 1–19). Springer. https://doi.org/10.1007/978-3-540-79228-4_1

[44]. Dyer, J. H., & Singh, H. (1998). The relational view: Cooperative strategy and sources of interorganizational competitive advantage. *Academy of Management Review*, *23*(4), 660–679. https://doi.org/10.5465/amr.1998.1255632

[45]. Ekanayake, E. M. A. C., Shen, G. Q. P., & Kumaraswamy, M. M. (2021a). Identifying supply chain capabilities of construction firms in industrialized construction. *Production Planning & Control*, *32*(4), 303–321. https://doi.org/10.1080/09537287.2020.1732494

[46]. Ekanayake, E. M. A. C., Shen, G. Q., Kumaraswamy, M. M., & Owusu, E. K. (2021b). Critical supply chain vulnerabilities affecting supply chain resilience of industrialized construction in Hong Kong. *Engineering, Construction and Architectural Management*, *28*(10), 3041–3059. https://doi.org/10.1108/ECAM-06-2020-0438

[47]. Elkan, C. (2001). The foundations of cost-sensitive learning. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI 2001)* (pp. 973–977). International Joint Conferences on Artificial Intelligence. https://www.ijcai.org/proceedings/2001/133

[48]. Fallah, A., Mokhtari, A., & Ozdaglar, A. (2020). Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)* (pp. 3557–3568). Curran Associates. https://proceedings.neurips.cc/paper/2020/hash/24389bfe4fe2eba8bf9aa9203a44cdad-Abstract.html

[49]. Fawcett, S. E., Magnan, G. M., & McCarter, M. W. (2008). Benefits, barriers, and bridges to effective supply chain management. *Supply Chain Management: An International Journal*, *13*(1), 35–48. https://doi.org/10.1108/13598540810850300

[50]. Fredrikson, M., Jha, S., & Ristenpart, T. (2015). Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS)* (pp. 1322–1333). ACM. https://doi.org/10.1145/2810103.2813677

[51]. Geiping, J., Bauermeister, H., Dröge, H., & Moeller, M. (2020). Inverting gradients: How easy is it to break privacy in federated learning? In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)* (pp. 16937–16947). Curran Associates. https://proceedings.neurips.cc/paper/2020/hash/c4ede56bbd98819ae6112b20ac6bf145-Abstract.html

[52]. Gentry, C. (2009). Fully homomorphic encryption using ideal lattices. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC 2009)* (pp. 169–178). ACM. https://doi.org/10.1145/1536414.1536440

[53]. Ghosh, A., Roughgarden, T., & Sundararajan, M. (2012). Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, *41*(6), 1673–1693. https://doi.org/10.1137/09076828X

[54]. Gondia, A., Siam, A., El-Dakhakhni, W., & Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, *146*(1), Article 04019085. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736

[55]. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML), PMLR 70* (pp. 1321–1330). PMLR. https://proceedings.mlr.press/v70/guo17a.html

[56]. Ha, A. Y., & Tong, S. (2008). Contracting and information sharing under supply chain competition. *Management Science*, *54*(4), 701–715. https://doi.org/10.1287/mnsc.1070.0795

[57]. Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning*, *77*(1), 103–123. https://doi.org/10.1007/s10994-009-5119-5

[58]. Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

[59]. Hao, M., Li, H., Luo, X., Xu, G., Yang, H., & Liu, S. (2020). Efficient and privacy-enhanced federated learning for industrial artificial intelligence. *IEEE Transactions on Industrial Informatics*, *16*(10), 6532–6542. https://doi.org/10.1109/TII.2019.2945367

[60]. Hsieh, K., Phanishayee, A., Mutlu, O., & Gibbons, P. (2020). The non-IID data quagmire of decentralized machine learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML), PMLR 119* (pp. 4387–4398). PMLR. https://proceedings.mlr.press/v119/hsieh20a.html

[61]. Hsu, J., Gaboardi, M., Haeberlen, A., Khanna, S., Narayan, A., Pierce, B. C., & Roth, A. (2014). Differential privacy: An economic method for choosing epsilon. In *2014 IEEE 27th Computer Security Foundations Symposium (CSF)* (pp. 398–410). IEEE. https://doi.org/10.1109/CSF.2014.35

[62]. Hu, R., Guo, Y., Li, H., Pei, Q., & Gong, Y. (2020). Personalized federated learning with differential privacy. *IEEE Internet of Things Journal*, *7*(10), 9530–9539. https://doi.org/10.1109/JIOT.2020.2991416

[63]. Hwang, B.-G., Zhao, X., & Ng, S. Y. (2013). Identifying the critical factors affecting schedule performance of public housing projects. *Habitat International*, *38*, 214–221. https://doi.org/10.1016/j.habitatint.2012.06.008

[64]. Hwang, S. (2011). Time series models for forecasting construction costs using time series indexes. *Journal of Construction Engineering and Management*, *137*(9), 656–662. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000350

[65]. Jagielski, M., Ullman, J., & Oprea, A. (2020). Auditing differentially private machine learning: How private is private SGD? In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)* (pp. 22205–22216). Curran Associates. https://proceedings.neurips.cc/paper/2020/hash/fc4ddc15f9f4b4b06ef7844d6bb53abf-Abstract.html

[66]. Jayaraman, B., & Evans, D. (2019). Evaluating differentially private machine learning in practice. In *Proceedings of the 28th USENIX Security Symposium* (pp. 1895–1912). USENIX Association. https://www.usenix.org/conference/usenixsecurity19/presentation/jayaraman

[67]. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., ... Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, *14*(1-2), 1–210. https://doi.org/10.1561/2200000083

[68]. Kairouz, P., Oh, S., & Viswanath, P. (2017). The composition theorem for differential privacy. *IEEE Transactions on Information Theory*, *63*(6), 4037–4049. https://doi.org/10.1109/TIT.2017.2685505

[69]. Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML), PMLR 119* (pp. 5132–5143). PMLR. https://proceedings.mlr.press/v119/karimireddy20a.html

[70]. Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., & Smith, A. (2011). What can we learn privately? *SIAM Journal on Computing*, *40*(3), 793–826. https://doi.org/10.1137/090756090

[71]. Kembro, J., Selviaridis, K., & Näslund, D. (2014). Theoretical perspectives on information sharing in supply chains: A systematic literature review and conceptual framework. *Supply Chain Management: An International Journal*, *19*(5/6), 609–625. https://doi.org/10.1108/SCM-12-2013-0460

[72]. Kerschbaum, F., Schröpfer, A., Zilli, A., Pibernik, R., Catrina, O., de Hoogh, S., Schoenmakers, B., Cimato, S., & Damiani, E. (2011). Secure collaborative supply-chain management. *Computer*, *44*(9), 38–43. https://doi.org/10.1109/MC.2011.224

[73]. Khan, L. U., Saad, W., Han, Z., Hossain, E., & Hong, C. S. (2021). Federated learning for Internet of Things: Recent advances, taxonomy, and open challenges. *IEEE Communications Surveys & Tutorials*, *23*(3), 1759–1799. https://doi.org/10.1109/COMST.2021.3090430

[74]. Kim, S.-Y., & Nguyen, V. T. (2018). A structural model for the impact of supply chain relationship traits on project performance in construction. *Production Planning & Control*, *29*(2), 170–183. https://doi.org/10.1080/09537287.2017.1398846

[75]. Kleijnen, J. P. C. (2005). An overview of the design and analysis of simulation experiments for sensitivity analysis. *European Journal of Operational Research*, *164*(2), 287–300. https://doi.org/10.1016/j.ejor.2004.02.005

[76]. Klein, R., & Rai, A. (2009). Interfirm strategic information flows in logistics supply chain relationships. *MIS Quarterly*, *33*(4), 735–762. https://doi.org/10.2307/20650325

[77]. Kshetri, N. (2018). 1 Blockchain's roles in meeting key supply chain management objectives. *International Journal of Information Management*, *39*, 80–89. https://doi.org/10.1016/j.ijinfomgt.2017.12.005

[78]. Kazi Mohammad Khalid, A., & Sumaiya Sadmeen, M. (2022). Geospatial Artificial Intelligence for Urban Flood Susceptibility and Drainage Resilience: A Prisma 2020 Systematic Review and Bibliometric Meta-Synthesis. *Review of Applied Science and Technology*, *12*(02), 01–51. https://doi.org/10.63125/epqdvy23

[79]. Le-Hoai, L., Lee, Y. D., & Lee, J. Y. (2008). Delay and cost overruns in Vietnam large construction projects: A comparison with other selected countries. *KSCE Journal of Civil Engineering*, *12*(6), 367–377. https://doi.org/10.1007/s12205-008-0367-7

[80]. Lee, H. L., So, K. C., & Tang, C. S. (2000). The value of information sharing in a two-level supply chain. *Management Science*, *46*(5), 626–643. https://doi.org/10.1287/mnsc.46.5.626.12047

[81]. Lee, J., & Clifton, C. (2011). How much is enough? Choosing ε for differential privacy. In *Information Security: ISC 2011, Lecture Notes in Computer Science 7001* (pp. 325–340). Springer. https://doi.org/10.1007/978-3-642-24861-0_22

[82]. Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

[83]. Li, L., Fan, Y., Tse, Y. K., & Lin, K.-Y. (2020a). A review of applications in federated learning. *Computers & Industrial Engineering*, *149*, Article 106854. https://doi.org/10.1016/j.cie.2020.106854

[84]. Li, L. (2002). Information sharing in a supply chain with horizontal competition. *Management Science*, *48*(9), 1196–1212. https://doi.org/10.1287/mnsc.48.9.1196.177

[85]. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020b). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50–60. https://doi.org/10.1109/MSP.2020.2975749

[86]. Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020c). Federated optimization in heterogeneous networks. In *Proceedings of Machine Learning and Systems 2 (MLSys 2020)* (pp. 429–450). mlsys.org. https://proceedings.mlsys.org/paper_files/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html

[87]. Li, X., Huang, K., Yang, W., Wang, S., & Zhang, Z. (2020d). On the convergence of FedAvg on non-IID data. In *International Conference on Learning Representations (ICLR 2020)*. OpenReview. https://openreview.net/forum?id=HJxNAnVtDS

[88]. Lindell, Y., & Pinkas, B. (2009). Secure multiparty computation for privacy-preserving data mining. *Journal of Privacy and Confidentiality*, *1*(1). https://doi.org/10.29012/jpc.v1i1.566

[89]. Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2020). Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. *IEEE Transactions on Industrial Informatics*, *16*(6), 4177–4186. https://doi.org/10.1109/TII.2019.2942190

[90]. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 54* (pp. 1273–1282). PMLR. https://proceedings.mlr.press/v54/mcmahan17a.html

[91]. McMahan, H. B., Ramage, D., Talwar, K., & Zhang, L. (2018). Learning differentially private recurrent language models. In *International Conference on Learning Representations (ICLR 2018)*. OpenReview. https://openreview.net/forum?id=BJ0hF1Z0b

[92]. Melis, L., Song, C., De Cristofaro, E., & Shmatikov, V. (2019). Exploiting unintended feature leakage in collaborative learning. In *2019 IEEE Symposium on Security and Privacy (SP)* (pp. 691–706). IEEE. https://doi.org/10.1109/SP.2019.00029

[93]. Meng, X. (2012). The effect of relationship management on project performance in construction. *International Journal of Project Management*, *30*(2), 188–198. https://doi.org/10.1016/j.ijproman.2011.04.002

[94]. Mironov, I. (2017). Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)* (pp. 263–275). IEEE. https://doi.org/10.1109/CSF.2017.11

[95]. Mohri, M., Sivek, G., & Suresh, A. T. (2019). Agnostic federated learning. In *Proceedings of the 36th International Conference on Machine Learning (ICML), PMLR 97* (pp. 4615–4625). PMLR. https://proceedings.mlr.press/v97/mohri19a.html

[96]. Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, *115*, 619–640. https://doi.org/10.1016/j.future.2020.10.007

[97]. Musarat, M. A., Alaloul, W. S., & Liew, M. S. (2021). Impact of inflation rate on construction projects budget: A review. *Ain Shams Engineering Journal*, *12*(1), 407–414. https://doi.org/10.1016/j.asej.2020.04.009

[98]. Mansura Akter, E., & Sumaiya, M. (2022). Antimicrobial Resistance Profiles of Acinetobacter Baumannii and Vibrio Cholerae in Hospital Wastewater and Adjacent Household Water Sources. *American Journal of Data Science and Analytics*, *3*(09), 01-55. https://doi.org/10.63125/gbervt14

[99]. Mashiat, N., & Md. Morshedul, I. (2022). Reducing Data Processing Latency in High Volume Telecommunications Networks Through Adaptive ETL and Distributed Computing. *American Journal of Scholarly Research and Innovation*, *12*(06), 01–36. https://doi.org/10.63125/b5xzay30

[100]. Md Imran, H., & Jannatul, F. (2022). Molecular and Precision Rehabilitation Molecular Mechanisms Linking Mitochondrial Dysfunction, Inflammation, and Muscle Degeneration with Functional Decline and Rehabilitation Response in Neuromuscular Disease. *American Journal of Interdisciplinary Studies*, *12*(08), 01-44. https://doi.org/10.63125/0ym5bh79

[101]. Md Maksudul, H., & Md Tazul Islam, C. (2022). Machine Learning-Based Cash-Flow Forecasting for Improving Financial Planning in U.S. Small and Mid-Sized Businesses. *International Journal of Scientific Interdisciplinary Research*, *12*(12), 01–38. https://doi.org/10.63125/w3pgjx97

[102]. Md. Morshedul, I., & Aminul Islam, M. (2022). A Quantitative Evaluation of Explainable Ensemble Learning for Probabilistic Risk Estimation and Decision Reliability in Enterprise Analytics. *American Journal of Data Science and Analytics*, *3*(02), 01-38. https://doi.org/10.63125/k22shk74Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (SP)* (pp. 111–125). IEEE. https://doi.org/10.1109/SP.2008.33

[103]. Nasr, M., Shokri, R., & Houmansadr, A. (2019). Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy (SP)* (pp. 739–753). IEEE. https://doi.org/10.1109/SP.2019.00065

[104]. Ng, S. T., Cheung, S. O., Skitmore, M., & Wong, T. C. Y. (2004). An integrated regression analysis and time series model for construction tender price index forecasting. *Construction Management and Economics*, *22*(5), 483–493. https://doi.org/10.1080/0144619042000202799

[105]. Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2021). Federated learning for Internet of Things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *23*(3), 1622–1658. https://doi.org/10.1109/COMST.2021.3075439

[106]. Paillier, P. (1999). Public-key cryptosystems based on composite degree residuosity classes. In *Advances in Cryptology: EUROCRYPT 1999, Lecture Notes in Computer Science 1592* (pp. 223–238). Springer. https://doi.org/10.1007/3-540-48910-X_16

[107]. Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Automation in Construction*, *122*, Article 103517. https://doi.org/10.1016/j.autcon.2020.103517

[108]. Pibernik, R., Zhang, Y., Kerschbaum, F., & Schröpfer, A. (2011). Secure collaborative supply chain planning and inverse optimization: The JELS model. *European Journal of Operational Research*, *208*(1), 75–85. https://doi.org/10.1016/j.ejor.2010.08.018

[109]. Polat, G., & Arditi, D. (2005). The JIT materials management system in developing countries. *Construction Management and Economics*, *23*(7), 697–712. https://doi.org/10.1080/01446190500041388

[110]. Rahman, M. A., Rahman, T., Laganière, R., Mohammed, N., & Wang, Y. (2018). Membership inference attack against differentially private deep learning model. *Transactions on Data Privacy*, *11*(1), 61–79. http://www.tdp.cat/issues16/tdp.a289a17.pdf

[111]. Reiter, J. P. (2005). Releasing multiply imputed, synthetic public use microdata: An illustration and empirical study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(1), 185–205. https://doi.org/10.1111/j.1467-985X.2004.00343.x

[112]. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, *3*(1), Article 119. https://doi.org/10.1038/s41746-020-00323-1

[113]. Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019). Blockchain technology and its relationships to sustainable supply chain management. *International Journal of Production Research*, *57*(7), 2117–2135. https://doi.org/10.1080/00207543.2018.1533261

[114]. Sablayrolles, A., Douze, M., Schmid, C., Ollivier, Y., & Jégou, H. (2019). White-box vs black-box: Bayes optimal strategies for membership inference. In *Proceedings of the 36th International Conference on Machine Learning (ICML), PMLR 97* (pp. 5558–5567). PMLR. https://proceedings.mlr.press/v97/sablayrolles19a.html

[115]. Said, H., & El-Rayes, K. (2011). Optimizing material procurement and storage on construction sites. *Journal of Construction Engineering and Management*, *137*(6), 421–431. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000307

[116]. Salem, A., Zhang, Y., Humbert, M., Berrang, P., Fritz, M., & Backes, M. (2019). ML-Leaks: Model and data independent membership inference attacks and defenses on machine learning models. In *Proceedings of the 26th Annual Network and Distributed System Security Symposium (NDSS 2019)*. Internet Society. https://doi.org/10.14722/ndss.2019.23119

[117]. Sambasivan, M., & Soon, Y. W. (2007). Causes and effects of delays in Malaysian construction industry. *International Journal of Project Management*, *25*(5), 517–526. https://doi.org/10.1016/j.ijproman.2006.11.007

[118]. Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation*, *7*(1), 12–24. https://doi.org/10.1057/jos.2012.20

[119]. Syed Feroz, A., & Md Mahamud Pervaz, P. (2022). Root-Cause Analysis and FMEA-Driven Reliability Improvement in Gas Production Facilities: A Multi-Year Empirical Study. *International Journal of Scientific Interdisciplinary Research*, *12*(10), 01–53. https://doi.org/10.63125/70msp878

[120]. Syed Feroz, A., & Shabbir, A. (2022). Finite Element-Based Thermomechanical Optimization of Pressure Components for Extending Industrial Boiler Service Life. *Journal of Sustainable Development and Policy*, *12*(12), 01-51. https://doi.org/10.63125/x24pht28

[121]. Segerstedt, A., & Olofsson, T. (2010). Supply chains in the construction industry. *Supply Chain Management: An International Journal*, *15*(5), 347–353. https://doi.org/10.1108/13598541011068260

[122]. Shahandashti, S. M., & Ashuri, B. (2013). Forecasting Engineering News-Record construction cost index using multivariate time series models. *Journal of Construction Engineering and Management*, *139*(9), 1237–1243. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000689

[123]. Shang, W., Ha, A. Y., & Tong, S. (2016). Information sharing in a supply chain with a common retailer. *Management Science*, *62*(1), 245–263. https://doi.org/10.1287/mnsc.2014.2127

[124]. Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)* (pp. 3–18). IEEE. https://doi.org/10.1109/SP.2017.41

[125]. Smith, V., Chiang, C.-K., Sanjabi, M., & Talwalkar, A. S. (2017). Federated multi-task learning. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (pp. 4424–4434). Curran Associates. https://papers.nips.cc/paper/2017/hash/6211080fa89981f66b1a0c9d55c61d0f-Abstract.html

[126]. Song, L., & Mittal, P. (2021). Systematic evaluation of privacy risks of machine learning models. In *Proceedings of the 30th USENIX Security Symposium* (pp. 2615–2632). USENIX Association. https://www.usenix.org/conference/usenixsecurity21/presentation/song

[127]. Song, S., Chaudhuri, K., & Sarwate, A. D. (2013). Stochastic gradient descent with differentially private updates. In *2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (pp. 245–248). IEEE. https://doi.org/10.1109/GlobalSIP.2013.6736861

[128]. Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(5), 557–570. https://doi.org/10.1142/S0218488502001648

[129]. Thunberg, M., & Persson, F. (2014). Using the SCOR model's performance measurements to improve construction logistics. *Production Planning & Control*, *25*(13-14), 1065–1078. https://doi.org/10.1080/09537287.2013.808836

[130]. Titus, S., & Bröchner, J. (2005). Managing information flow in construction supply chains. *Construction Innovation*, *5*(2), 71–82. https://doi.org/10.1108/14714170510815186

[131]. Triastcyn, A., & Faltings, B. (2019). Federated learning with Bayesian differential privacy. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 2587–2596). IEEE. https://doi.org/10.1109/BigData47090.2019.9005465

[132]. Truex, S., Baracaldo, N., Anwar, A., Steinke, T., Ludwig, H., Zhang, R., & Zhou, Y. (2019). A hybrid approach to privacy-preserving federated learning. In *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security (AISec 2019)* (pp. 1–11). ACM. https://doi.org/10.1145/3338501.3357370

[133]. Truex, S., Liu, L., Chow, K.-H., Gürsoy, M. E., & Wei, W. (2020). LDP-Fed: Federated learning with local differential privacy. In *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking (EdgeSys 2020)* (pp. 61–66). ACM. https://doi.org/10.1145/3378679.3394533

[134]. Vrijhoef, R., & Koskela, L. (2000). The four roles of supply chain management in construction. *European Journal of Purchasing & Supply Management*, *6*(3-4), 169–178. https://doi.org/10.1016/S0969-7012(00)00013-7

[135]. Wang, J., Liu, Q., Liang, H., Joshi, G., & Poor, H. V. (2020). Tackling the objective inconsistency problem in heterogeneous federated optimization. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)* (pp. 7611–7623). Curran Associates. https://proceedings.neurips.cc/paper/2020/hash/564127c03caab942e503ee6f810f54fd-Abstract.html

[136]. Wang, Y.-X., Balle, B., & Kasiviswanathan, S. P. (2019). Subsampled Rényi differential privacy and analytical moments accountant. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 89* (pp. 1226–1235). PMLR. https://proceedings.mlr.press/v89/wang19b.html

[137]. Wedawatta, G., Ingirige, B., & Amaratunga, D. (2010). Building up resilience of construction sector SMEs and their supply chains to extreme weather events. *International Journal of Strategic Property Management*, *14*(4), 362–375. https://doi.org/10.3846/ijspm.2010.27

[138]. Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., Jin, S., Quek, T. Q. S., & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Transactions on Information Forensics and Security*, *15*, 3454–3469. https://doi.org/10.1109/TIFS.2020.2988575

[139]. Xue, X., Wang, Y., Shen, Q., & Yu, X. (2007). Coordination mechanisms for construction supply chain management in the Internet environment. *International Journal of Project Management*, *25*(2), 150–157. https://doi.org/10.1016/j.ijproman.2006.09.006

[140]. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019a). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology*, *10*(2), 1–19. https://doi.org/10.1145/3298981

[141]. Yang, W., Zhang, Y., Ye, K., Li, L., & Xu, C.-Z. (2019b). FFD: A federated learning based method for credit card fraud detection. In *Big Data: BigData 2019, Lecture Notes in Computer Science 11514* (pp. 18–32). Springer. https://doi.org/10.1007/978-3-030-23551-2_2

[142]. Yao, A. C.-C. (1986). How to generate and exchange secrets. In *27th Annual Symposium on Foundations of Computer Science (SFCS 1986)* (pp. 162–167). IEEE. https://doi.org/10.1109/SFCS.1986.25

[143]. Yaseen, Z. M., Ali, Z. H., Salih, S. Q., & Al-Ansari, N. (2020). Prediction of risk delay in construction projects using a hybrid artificial intelligence model. *Sustainability*, *12*(4), Article 1514. https://doi.org/10.3390/su12041514

[144]. Yeom, S., Giacomelli, I., Fredrikson, M., & Jha, S. (2018). Privacy risk in machine learning: Analyzing the connection to overfitting. In *2018 IEEE 31st Computer Security Foundations Symposium (CSF)* (pp. 268–282). IEEE. https://doi.org/10.1109/CSF.2018.00027

[145]. Zainal Abidin, N. A., & Ingirige, B. (2018). The dynamics of vulnerabilities and capabilities in improving resilience within Malaysian construction supply chain. *Construction Innovation*, *18*(4), 412–432. https://doi.org/10.1108/CI-09-2017-0079

[146]. Zhao, Y., Zhao, J., Yang, M., Wang, T., Wang, N., Lyu, L., Niyato, D., & Lam, K.-Y. (2021). Local differential privacy-based federated learning for Internet of Things. *IEEE Internet of Things Journal*, *8*(11), 8836–8853. https://doi.org/10.1109/JIOT.2020.3037194

[147]. Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)* (pp. 14774–14784). Curran Associates. https://papers.nips.cc/paper/2019/hash/60a6c4002cc7b29142def8871531281a-Abstract.html