Optimizing Alpha through Better Information Workflows

Chito Jovellanos

Chito Jovellanos
is president of *forward look*, inc., in Boston, MA.
cjovellanos@forwardlook.com

KEY FINDINGS

- For portfolio managers, implementation shortfalls stemming from suboptimal information workflows can be systematically mitigated to maximize the expression of alpha inherent in an investment strategy.
- For operational managers, re-aligning investment operations workflows to deliver differentiated support across individual strategies can significantly improve portfolio performance.
- For data scientists, graph models provide fresh evidence of the quantifiable impacts of data interoperability and information parity on the efficiency of enterprise information flows.

ABSTRACT

This study demonstrates how investment managers can identify and resolve suboptimal operational workflows that diminish an investment strategy's attainable alpha on the order of 24-242 basis points (annualized, gross of fees). A portfolio's potential alpha can be best realized by addressing these portfolio implementation shortfalls through systematic improvements in the data schemas that drive the information exchanges among parties regarding transactions, holdings, and valuations. Using graph modelling techniques, we identify and resolve information workflow inefficiencies that occur both within and external to the investment firm.

TOPICS

*Portfolio management, portfolio construction, equity portfolio management, fixed-income portfolio management, quantitative methods, statistical methods**

In Jovellanos (2011), we described how a portfolio's projected alpha was seemingly muted by inefficiencies in investment operations workflows. Based on that study of 61 portfolios from 14 asset managers across varied investment strategies, the curtailment in returns attributable to portfolio implementation shortfalls (Perold 1988) spanned 51 to 242 bps (annualized, gross of fees).

Our 2011 study also discerned that these shortfalls can be offset by applying targeted improvements to information workflows. The ability to better retain alpha arose, for example, from simply reducing errors (e.g., failed trades due to incorrect security identifiers) all the way to mitigating a range of opportunity costs (e.g., timely "locates" for borrowing and lending initiatives or up-to-date foreign cash availability enabling greater latitude for new buys).

*All articles are now categorized by topics and subtopics. View at PM-Research.com.

Investment operations is a complex endeavor in data management, which can involve screeners that rely on shifting ESG taxonomies, optimizers that require quantifiable constraints, liquidity seeking algos that rely on data embedded in IOIs and conditional orders, risk management systems that depend on fixed income analytics, or the ISDA CDM (Common Domain Model) protocol that underpins "smart contracts" for derivatives settlement. We therefore hypothesized that maximal expression of alpha arose from more efficient operational workflows driven by better quality of information exchanged between operational units within a firm (e.g., Research to Trading) as well as their external counterparties (e.g., Trading to Prime Broker). We characterized the quality of information via the *Data Operability Threshold* (DOT) metric[1] that was inversely correlated to performance (see Appendix in Jovellanos [2011] for a detailed description).

Since that 2011 study, we obtained access to an additional 52 portfolios from 11 managers across a range of investment strategies, which allowed us to further validate the robustness of the DOT metric. From these, we selected and analyzed 38 portfolios from 8 managers. These sample portfolios had to be amenable to deep inspection and meet the selection criteria consistent with our prior initiative's framework (see Methodology section in Jovellanos 2011). Exhibit 1 summarizes the characteristics of the latest portfolios we examined. Exhibit 2 reiterates the strong inverse correlation between the DOT value and observed performance gains in these latest samples, with $r^2 = 0.85$, akin to the 2011 study.

In addition to broadening the span of strategies and portfolios examined, this newer data set also gave us the opportunity to incorporate novel analytic techniques to amplify the utility of the DOT metric. In particular, graph modelling, which is widely used not just in social network analysis but also in a range of industries (Hegeman and Iosup 2018), allowed portfolio managers we worked with to:

1. Identify the specific segments across a firm's entire information workflow that diminished a strategy and its associated portfolio's potential alpha; and
2. Envision alternative operational workflows that could help mitigate alpha dissipation due to sub-optimal data exchanges regarding transactions, holdings, and valuations between operating units, both within a firm and with its external counterparties.

We highlight numerous examples that illuminate our key insights regarding how best to tune information workflows to optimize the expression of alpha from an investment strategy.

METHODOLOGY

We envisioned a manager who wishes to deploy a specific investment thesis, and poses the question of how portfolio implementation workflows can be structured to maximize the realized return available from the strategy. We examined investment operations atypically—through the lens of a portfolio manager, rather than the more traditional middle or back-office perspective. We then applied graph modelling tech-

[1] In brief, DOT quantified the accuracy and usability of the syntax ("format") and semantics ("meaning") of these information exchanges. DOT also incorporated a Bayesian component that weighted the firm's operating history with the securities transacted. DOT values were inversely correlated with portfolio performance. High DOT values (closer to 1.0) were symptomatic of detrimental "hot-spots" in the firm's investment operations workflow, whereas lower DOT values (closer to 0.0) were indicative of efficient information flows.

EXHIBIT 1

Summary of Sampled Portfolios (January 2010–November 2018)

	Operational Facet	Obs.Period (mm-yy) Start	End	Months	Strategy	Securities	Vehicle	Base Curr.	Portfolio Market Value (MM)	Performance (bp) Start	End	Gain	Remediation	DOT
PORTFOLIO ANALYTICS														
1	Market Data	Feb-18	Aug-18	7	dev. mkts large cap	eq, cash, swaps	sep.acct	USD	982	3	48	45	scrub multiple sources	0.61
2		Nov-17	May-18	7	global small cap	eq, cash, swaps	sep.acct	USD	545	54	104	50	scrub multiple sources	0.49
3		Jan-13	Nov-13	11	asset alloc	eq, bond, opt, fut, cash, swap	sep.acct	USD	388	–8	28	36	scrub multiple sources	0.75
4	Macro Data	Mar-18	Nov-18	9	risk parity	eq,bond,opt, fut, cash, swap	sep.acct	USD	1,393	10	77	67	global industry data	0.47
5		Dec-11	Jan-13	14	full discretion	eq, bond, cash	sep.acct	USD	1,102	2	88	86	global inflation data	0.38
6	'Alt' Data (for Machine Learning)	Jan-17	Feb-18	14	liability driven	bond, opt, fut, cash, swap	sep.acct	USD	1,310	15	67	52	third-party data review	0.56
7		Mar-16	Nov-16	9	dev. mkts small cap	eq, cash, swaps	sep.acct	USD	756	21	187	166	momentum indicators	0.12
8		Feb-16	Oct-16	9	US small cap growth	eq, cash	sep.acct	USD	598	52	143	91	geographic data	0.25
9		Jan-14	Dec-14	12	emerging	eq, cash, swaps	sep.acct	USD	390	–22	25	47	interest rate data	0.54
10		Mar-14	Oct-14	8	US large cap value	eq, fut, cash, swaps	sep.acct	USD	675	41	127	86	sentiment indicators	0.43
11		Feb-13	Dec-13	11	global credit	bond, fut, cash, swaps	sep.acct	EUR	748	18	64	46	market transition indicators	0.58
12		Sep-12	Feb-13	6	global credit	bond, cash, swaps	sep.acct	USD	588	–22	87	109	sentiment indicators	0.21
13	Risk Model	Mar-16	Nov-16	9	emerging mkts	eq, cash, swaps	fund	USD	862	3	75	72	benchmark management	0.33
14		Aug-14	Mar-15	8	dev. mkts growth	eq, bond, fut, cash, swaps	sep.acct	USD	931	159	39	120	factor management	0.09
15		Apr-12	Oct-12	7	US small cap	eq, fut, cash, swaps	sep.acct	USD	102	107	49	58	model metadata	0.5
COMPLIANCE														
16	Pre-Trade Checks	Nov-17	Jul-18	9	US mid-small cap	eq, fut, cash, swaps	sep.acct	USD	742	43	118	75	corp actions mgmt	0.36
17		Mar-16	Jan-17	11	US large cap	eq, fut, cash	sep.acct	USD	869	31	63	32	corp actions mgmt	0.57
18		Mar-16	Oct-16	8	dev. mkts credit	bond, cash, swaps	sep.acct	USD	438	–8	49	47	exposure calcs	0.49
19		Jul-16	Oct-16	4	US small cap growth	eq, cash	sep.acct	USD	83	33	161	128	corp action smgmt	0.14
20		Jan-15	Dec-15	12	US small cap value	eq, cash	sep.acct	USD	52	93	187	94	exposure calcs	0.21
21		May-15	Dec-15	8	liability driven	eq, bond, opt, fut, cash, swap	sep.acct	USD	1,554	19	43	24	corp action smgmt	0.66
22		Apr-14	Nov-14	8	asset alloc	eq, bond, opt, fut, cash, swap	fund	USD	972	63	88	25	exposure calcs	0.79
23		Mar-13	Nov-13	9	US mid cap	eq, fut, cash, swaps	sep.acct	USD	697	43	95	52	exposure calcs	0.51
24		Apr-11	Feb-12	11	global credit	bond, fut, cash, swaps	sep.acct	USD	807	44	111	67	margin checks	0.42
25	Post-Trade Checks	Jul-15	Mar-16	9	risk parity	eq, bond, optfut, cash, swaps	sep.acct	USD	706	82	262	180	exposure calcs	0.08
26		Mar-11	Dec-11	10	US credit	bond, fut, cash, swaps	sep.acct	USD	317	1	33	32	parent level checks	0.63
27		Jan-10	Sep-10	9	global bond	bond, fut, cash, swaps	fund	USD	688	37	78	41	firm level checks	0.58

(continued)

EXHIBIT 1 *(continued)*

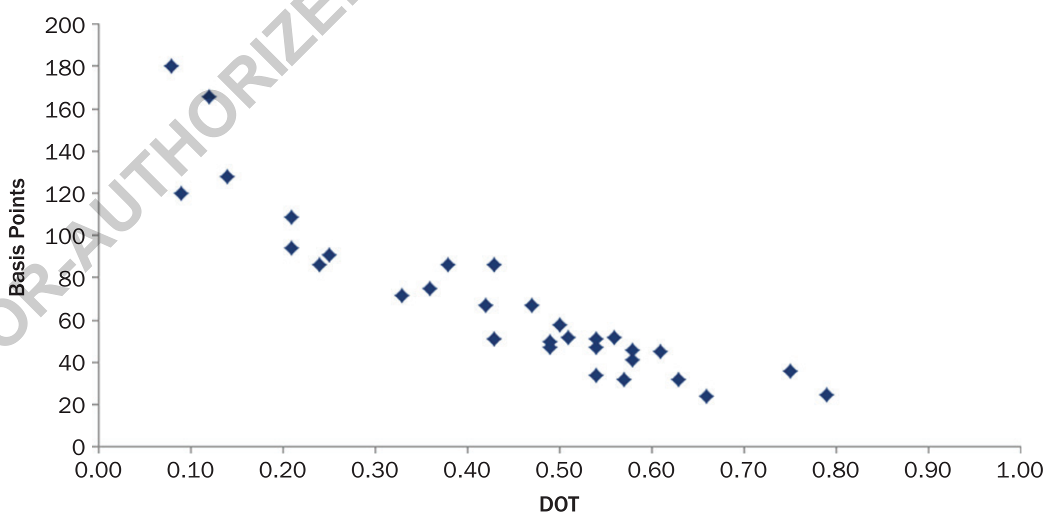
Summary of Sampled Portfolios (January 2010–November 2018)

	Operational Facet	Obs.Period (mm-yy) Start	End	Months	Strategy	Securities	Vehicle	Base Curr.	Portfolio Market Value (MM)	Performance (bp) Start	End	Gain	Remediation	DOT
TRADING														
28	Execution	Nov-17	Nov-18	13	min volatility	eq, bond, fut, cash	sep.acct	USD	98	51	102	51	IOI management	0.43
29		Aug-17	Mar-18	8	US small cap growth	eq, cash	sep.acct	USD	126	13	64	51	broker interfaces	0.54
30		Mar-15	Dec-15	10	asset alloc	eq, bond, fut, cash, swaps	sep.acct	USD	264	19	53	34	algo implementation	0.54
31		May-13	Dec-13	8	global credit	bond, cash, swaps	sep.acct	USD	355	22	93	86	OMS workflow	0.24
32		Feb-13	Nov-13	10	emerging mkts	eq, cash, swaps	sep.acct	USD	512	99	155	56	IOI management	0.36
33		Mar-12	Dec-12	10	balanced	eq, bond, fut, cash, swaps	fund	USD	1,470	63	97	34	broker interfaces	0.69
34		Feb-11	Oct-11	9	US fixed income	bond, fut, cash, swaps	sep.acct	USD	867	–5	48	53	broker interfaces	0.32
35		Jan-10	Jul-10	7	asset alloc	eq, bond, opt, fut, cash, swap	sep.acct	EUR	505	31	85	54	merge fills from MTF's	0.33
36	Shorting	Jul-17	Aug-18	14	global opp's	eq, bond, fut, cash, swap	sep.acct	USD	1,310	95	149	54	prime broker interface	0.67
37		Feb-13	Dec-13	11	US mid cap	eq, cash, swaps	fund	USD	628	–10	41	51	prime broker interface	0.59
SETTLEMENTS														
38	FX	Oct-11	Jul-12	10	dev. mkts equity	eq, cash	sep.acct	USD	208	21	79	58	algo implementation	0.22

NOTES: Grouped by Investment Operations facet. Portfolio performance (in basis points, relative to their individual benchmarks: annualized, gross of fees) reported at start and end of observation period. Portfolio value reflects month-end AuM reported at the midpoint of the observation period. DOT values shown concurrent with performance reported at the end of the observation period.

EXHIBIT 2

Aggregate Performance Improvement vs. Concurrent DOT Values for the Portfolios in Exhibit 1 (in bps, relative to the applicable benchmark, annualized, gross of fees)



niques[2] to evaluate the efficacy of day-to-day operational activities undertaken by the entire firm and its external counterparties to enable the fullest expression of a strategy's alpha across one or more portfolios.

Operational Framework

Investment management businesses all share a common operating structure: linked interacting entities, both within a firm (such as portfolio managers, traders, compliance, settlements, etc.), and with external counterparties (such as clients, prime brokers, custodians, data providers, regulators, etc.). We captured the generality of this fundamental structure in a reference graph model (Exhibit 3) that served as the starting template for all subsequent firm-specific models.

Each "node" in the graph represents an operational unit (e.g., portfolio managers), and each path between nodes ("edges" in graph modelling terms) represents the data flowing to and from those nodes (e.g., proposed trades between compliance and portfolio managers). In this study, individual paths embody information flows about the market value of a particular asset type (e.g., non-agency CMOs) directed between nodes over some unit of time (e.g., at every book-of-record checkpoint[3]).

Given the importance of indexes to investment processes, our model incorporates a distinct node for index providers in order to capture their influence in setting and measuring investment strategies (Robertson 2019). Distinct nodes also represent third-party research (given its heightened regulatory profile as a consequence of MiFID II), as well as reference data sources that include traditional market data vendors, entities such as loan servicers for asset-backed markets, and novel "nowcasting" providers that supply near-time metrics of retail shopping activity, electricity consumption, and so on.

Models versus Reality

In practice, this generic representation of investment management workflows must be elaborated uniquely for each investment firm studied in order to characterize:

1. Specific strategies the firm supports (e.g., global credit, LDI, tax managed)
2. Information sources on which it relies (e.g., bulk shipping data, ESG metrics)
3. Instruments it trades (e.g., convertibles, CLOs, options on futures)
4. Investment constraints under which it operates (e.g., credit ratings, RegS/144a, margining)
5. Proprietary advantages of the firm (e.g., curated sentiment data, private markets access)
6. Limits of its operating units (e.g., minimal experience with borrowing and lending).

Exhibit 4 is a visual representation of the graph model for an asset manager that was covered in our study. The model shown is specific to the "global opportunities" strategy that was implemented in one underlying portfolio (Exhibit 1: portfolio #36).

[2] The underlying graph network data was maintained and profiled using the R package *igraph* (https://igraph.org/r/). Network visualizations and analysis were enabled via *Gephi* (https://gephi.org/) and *Hive* (https://hiveplot.com/).

[3] The most pragmatic definition of an Investment Book of Record comes from Blackwell (2014): it delivers the current best available view of data suitable for investment decision-making, including:

a) the current status and forward projections of:

- Portfolio holdings, including transactions and security lifecycle events, and their statuses
- Cash positions, including transactions and their statuses

b) reference data and derived analytics supporting the investment decision making process.

EXHIBIT 3

Template—Reference Graph Model: Institutional Investment Workflows

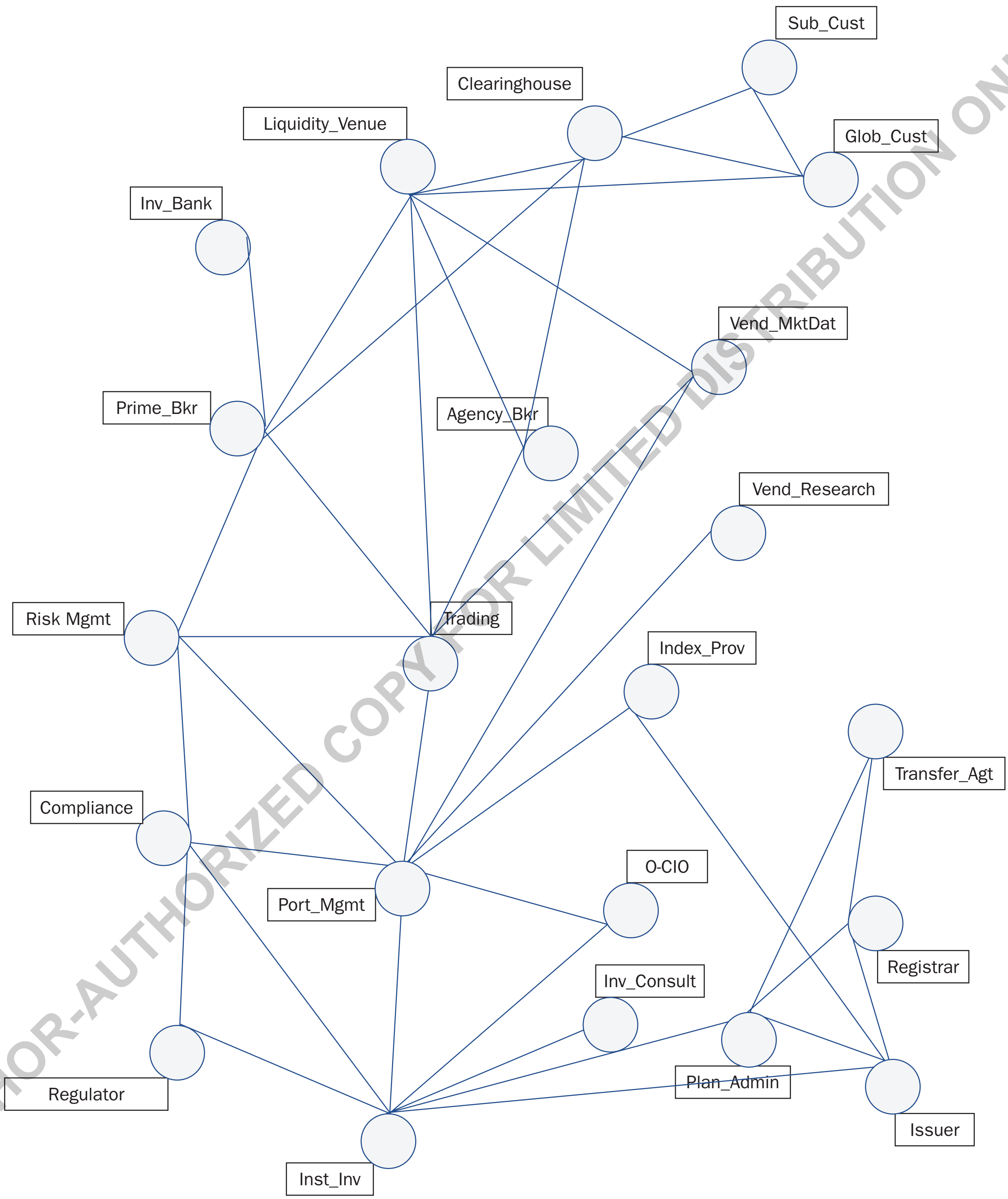
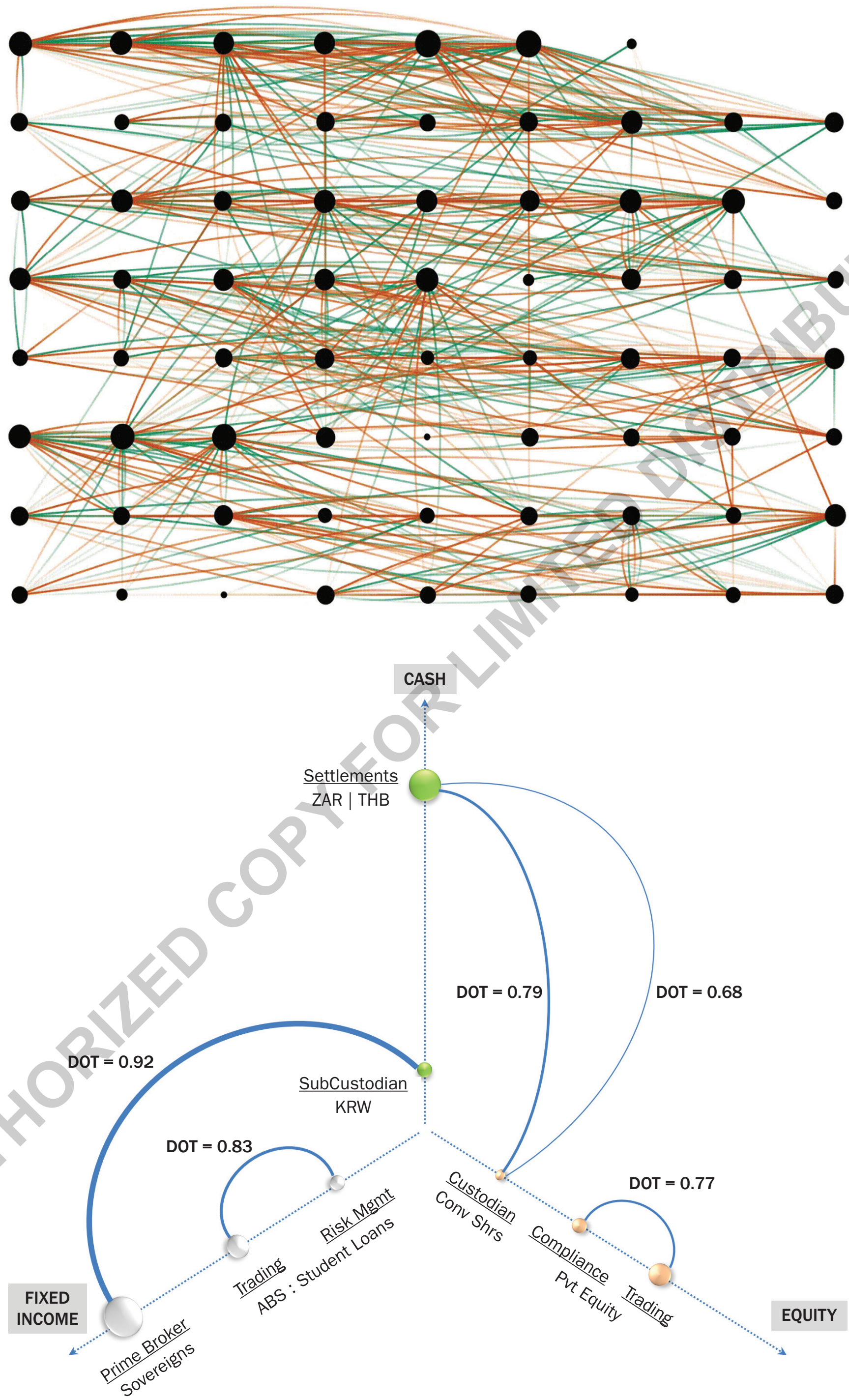


EXHIBIT 4

Example: Firm-Specific Graph Model



UPPER PANEL: Pro-forma visualization of a small region of a firm's graph model. Each layer reflects the operational flows supporting a specific asset type (e.g., CMBS or cash). Functional units or "nodes" (e.g., Trading, Compliance) are represented in each layer. The larger the node, the more "central" (important) it is. The interconnections or "edges" embody the information flows representing market values of the assets in the portfolio. The efficiency of these flows is weighted via the DOT metric.

Note that the more complex the model, the less readable this pro-forma graph model visualization becomes.

LOWER PANEL: To ensure that "hot-spots" can be readily discerned, we recast the pro-forma graph model visualization into a hive plot. For illustrative purposes, only edges whose DOT value exceeds 0.60 are shown. Each edge's weight (portrayed via the thickness of the path) reflects the market value of the security type being handled. The firm's functional units (nodes) are arrayed according to importance ("centrality") along each of three axes that represent the major asset classes (cash, fixed income, equity).

Note the multi-layered structure. Each plane of the model captures the data flows for a specific asset type (e.g., Fixed Income: ABS: Student Loans). Each plane then interacts with other planes (e.g., as exposures to different asset classes are revised over time), resulting in a composite picture of the operational workflow that services a portfolio in support of an investment strategy.

For fixed income instruments and derivatives, security types are based on the Bloomberg Barclays Level 4 Classification scheme. For equity-based instruments, we tracked them as one of common, preferred, options, futures, swaps, rights, warrants, ADRs/GDRs, ETFs, or hedge funds (including partnerships).

In our original 2011 study, we examined virtually every transaction, message, and file exchanged between different parties (both within and outside of the firm) in order to profile the information flows over that firm's network. Over time, we realized that the essential information we needed to capture was already called out in investment management agreements and guidelines. In effect, investment constraints define the scope of a firm's mandate and capture the essential boundaries that bracket the manager's strategies and its associated portfolios. As a result, these constraints are sufficient for determining the graph model structure that represents the essential workflows that support a specific strategy and its associated portfolio(s). The resulting graph model was corroborated using the compliance checks (both pre- and post-trade) built into a firm's order and risk management systems as well as any documented manual processes. Every compliance rule had to be able to map back to discrete nodes and paths in the graph model, and vice-versa.

Baselining Information Quality

We used graph models to visualize and analyze the firms' existing information flows. Based on actual events culled from a firm's various systems (trading logs, accounting records, compliance entries, etc.), we simulated the sequence of operational activities that were material to expressing an investment strategy across its underlying portfolio(s) over time, such as cash inflows, initial buys, transfers in-kind, rebalancings, amending errors, block allocations, posting corporate actions, tracking income and pay-downs, and so on. Using discrete simulation software,[4] we pulsed these events through the graph model's nodes and paths using the time steps dictated by the firm's intra-day and/or end-of-day information cycle.[5] The simulation enabled us to trace network-wide information flows by "following the money," seeing how data about transactions, positions, and valuations (for cash and all security types) were propagated through the firm's information web—from research to trading, through external servicers, and on to their eventual disposition from the firm's portfolio(s). At a minimum, we captured end-of-day data for six contiguous months. The use of actual historical data from the firm's systems provided us with a baseline ("as-is") view that showed how the existing workflow utilized data and conditioned information quality for the strategy expressed in a portfolio.

The simulations recorded the changes in DOT values across all the network paths over each intra-day run cycle. At each time step, we calculated a DOT value for every path between adjacent interacting nodes. Recall that DOT quantifies the quality of information transfer from source to target node[s] and is associated with the information flows specific to an asset type. The DOT value for each path then

[4] We used the R package "simmer" (https://r-simmer.org/).

[5] Interestingly, we never encountered a firm with a comprehensive implementation of an Investment Book of Record. Typically, there were at most three start-of-day and three end-of-day observations (for the Asian, American, and European markets). Only two managers in this study captured a mid-day snapshot of pivotal operational activities.

was assigned a weight based on the market value of the asset type relative to the total market value of the portfolio.

We also derived a top-level (model-wide) DOT value at each time step as the weighted average of the individual path-specific DOT values, and used it as a proxy for the overall health of the firm's information workflow for the strategy and its associated portfolios.

While characterizing the baseline view, we also examined structural features of the firm-specific graph network and evaluated their impact on information flows. For example, notable groups of nodes or particular pathways encompassing multiple edges may be central to enhancing or impeding efficient information flow for a particular asset type. Likewise, the extent of node clustering (including hierarchical clusters), the number of connections between nodes, the centrality of nodes and edges, and the inter-dependencies across network layers are known to influence how information propagates across a network.

Establishing an Optimal Workflow

We used the DOT values generated by the simulations from the baseline view to identify the workflow nodes and paths where information quality was impaired. DOT values closer to 1.0 (see Exhibit 4 for examples) were indicative of high obstacles to information transfer, and were often, but not always, marked by high levels of reconciliation activity—both manual and automated.

To remedy these workflow hot-spots and resolve the underlying inefficiencies, there are three alternatives available in the context of a firm's graph model:

1. *Reducing edges* by consolidating flows between nodes. A typical example might involve managing credit risk more efficiently by aggregating flows associated with high-yield instruments (bank loans, subprime mortgages, etc.) separately from flows stemming from investment-grade securities.
2. *Reducing nodes* by disintermediating unnecessary participants or procedures in the workflow. An extreme example would be an institutional investor repatriating strategies in-house, thereby eliminating nodes (in addition to flows) involving a third-party investment manager. Other examples are consolidation of disparate accounting systems, or rationalizing overlapping market data and research vendor feeds.
3. *Reinforcing nodes* by focusing on the efficiency with which nodes process and communicate data with other linked nodes. This could be expressed in a range of overt remedial actions such as re-designed functional processes (e.g., automate and streamline corporate actions data distribution for voluntary events, or improve the timeliness of private asset valuations) or infrastructure upgrades (e.g., adopt fund accounting systems that offer better handling for hedge funds and partnerships). However, the essence of node reinforcement is the adoption of more efficient data schemas used within the node to optimize data extraction, loading, and consumption across adjacent nodes (such as upgrading to an FPML message version that better accommodates CLOs for Japanese investors' portfolios, or harmonizing FIX message usage with counterparties).

These alternative approaches to workflow optimization are not mutually exclusive and may in fact be complementary. However, for the purposes of this study, we only evaluated portfolios whose remediation approach exploited only one of these options so as not to confound the outcomes, much like our initial 2011 study.

The specific changes applied to nodes and edges require a robust understanding of the investment strategy being supported to ensure the potential modifications are sensible. Therefore, participation by the portfolio manager was crucial. Candidates for workflow changes were rooted in addressing the impairments identified in the baseline analysis. Although occasionally compelling, we eschewed anecdotal evidence about likely sources of implementation shortfalls, and managers' intuition about where the most efficiencies could be gained. We relied on such observations only to corroborate the analysis stemming from the baseline workflow analysis.

Measuring Potential Gains

The measurable impact of potential workflow changes to portfolio performance can be gauged by evaluating a revised DOT profile, and correlating the new DOT values to projected gains (in bps) based on the linear relationship per Exhibit 2.[6] The underlying equation is:

$$\mathbf{Y} = (-160.39 * \mathbf{X}) + 136.46$$

where Y = gain in basis points
X = DOT value
either at the path-level for localized analysis, or top-level for an aggregate network-wide measure.

Obtaining a revised DOT profile involves replaying the same event log that was originally used to elicit the initial DOT profile associated with the baseline ("as-is") graph model. We redirected select log entries as needed to reflect where nodes or paths were either eliminated or consolidated. For example, if a subcustodian was replaced by a global custodian, we directed the original flows to the latter—resulting in revised path-level DOT metrics. Where node reinforcement was undertaken, affected nodes emitted different DOT metrics given the implicit changes to the underlying data schema(s). As with the original baseline graph model, we pulsed the updated events through the revised graph model's nodes and paths using the same time steps contingent on the firm's intra-day or end-of-day information cycle to obtain updated DOT values.

Subsequent comparison of the path-level DOT outputs between the baseline versus optimized model scenarios highlighted the nodes and paths in the investment operations network that were material contributors to the decay in performance—and more importantly, how the improved DOT profile translated into maximizing attainable alpha. Moreover, since there were a number of viable changes to the baseline network's nodes and paths, the simulation also gave us the ability to project how a range of alternative operational workflows could help mitigate specific features of the implementation shortfalls that were initially identified.

[6] The correlations seem specific to a prevailing market regime. In our 2011 paper, we reported that 51-242 bps in inherent alpha (annualized, gross of fees) could be preserved through more efficient operational workflows. The current study reports lower values of 24-180 bps in retained alpha. The diminished upside is expected, given the materially lower return environment following the 2008 Lehman-induced crisis. Distance metrics suggest two distinct clusters between the two market regimes, with a centroid of 98 bps in the current study versus 127 bps in our 2011 study. The comparable equation for the 2011 study dataset is $\mathbf{Y} = (-201.52 * \mathbf{X}) + 235.04$.

OBSERVATIONS

From the portfolios listed in Exhibit 1, we highlight exemplars that best illuminate how material improvements in the expression of a strategy's innate alpha were enabled by optimizing portfolio implementation workflows through edge reduction, node reduction, or node reinforcement.

Reducing Edges

A global developed markets small-cap implementation (Exhibit 1: portfolio #7) sought to exceed its benchmark (a custom blend of the MSCI All Country World Small Cap Index and the S&P SmallCap 600) and meet its goal of long-term capital appreciation. The investment team surmised that the attainable alpha from the small-cap sector is materially conditioned by rapidly evolving business drivers that animate fundamental company performance. To that end, they acquired unique information sources ("alt-data") that would better inform them in near-time about potential impacts to the businesses of their portfolio companies (e.g., remote sensing profiles to measure coastal water quality and wave activity, and how these might modify farmed fish production levels in specific grow-out regions).

The analysis of their baseline workflows indicated hot-spots centered around the use of these alt-data sources (DOT > 0.72). Over time, this institutional investor had accrued 11 alt-data sources, each providing information consumed in isolation. Taking a step back and viewing potential synergies in data management across their portfolios, they opted to partition these 11 sources into four factors that they believed more broadly influenced the performance of their portfolio companies (e.g., coastal wave activity was material to investments not just in aquaculture, but also shipping, port operations, and recreation). They then modified the information workflows (and the associated graph model) to manage these four flows explicitly. The DOT profile of the revised graph model forecast a 149 bp potential gain, with the actual gains coming to 166 bp.

Across all the portfolios examined in this current study, edge reduction initiatives yielded anywhere from 34-166 bp (annualized, gross of fees).

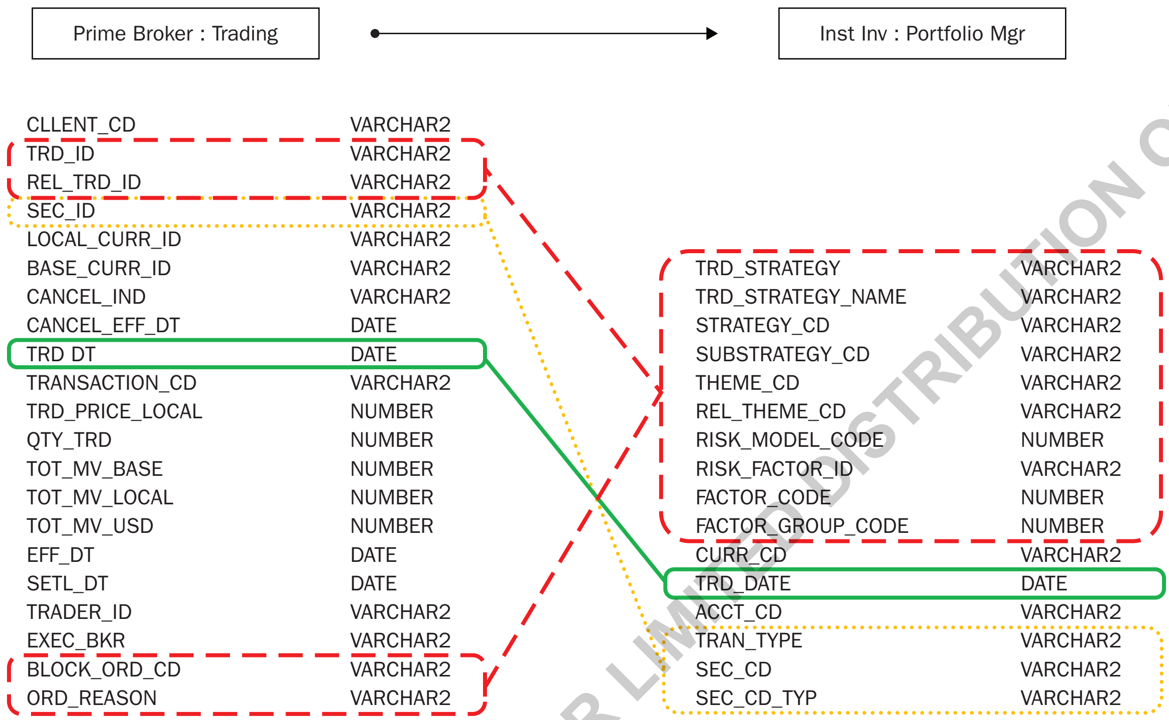
Reducing Nodes

A US small-cap implementation (Exhibit 1: portfolio #19) was part of a sovereign wealth fund's initiative to broaden and diversify its sources of returns. This growth-oriented portfolio was benchmarked against a custom blend of the MSCI U.S. Small Cap 1750 and the Russell 2000.

A few months after inception, inefficiencies in corporate actions processes became evident from the analysis of their baseline workflows (DOT > 0.61). The fund was being serviced by a network of sub-custodians with which it had a long-standing history; moreover, services across the different banks had been apportioned based on their reputed familiarity with specific companies and industries. An operations audit confirmed that the portfolio was subject to late, missed, and/or erroneous corporate actions—all attributable in one form or another to the existing sub-custodians. To remedy the situation, the fund directed a different third-party servicer to assume responsibility for all corporate actions processing. This workflow change resulted in the reduction of nodes (functional parties) in the firm's graph model that pointed to a projected gain of 131 bp. The actual gains came to 128 bp.

Across all the portfolios examined in this current study, node reduction initiatives yielded anywhere from 24-128 bp (annualized, gross of fees).

EXHIBIT 5
Data Schema Example



NOTES: Segment of data schemas shown for originating node (Prime Broker: Trading) and recipient node (Institutional Investor: Portfolio Manager). Solid lines bracket data elements exhibiting high information fidelity in terms of both semantics (trade date) and syntax (date format) of the data. Dotted lines bracket data elements with a potential risk of misrepresentation from source to target node (security identifier relies on lookup tables to resolve). Dashed lines mark wide disparities in source to target schema elements where additional software and reference tables external to the database are required to resolve the mappings. The DOT value (0.88) for this specific segment of the graph model flow confirms poor data interoperability.

Reinforcing Nodes

A global multi-asset fund employing levered risk-parity (Exhibit 1: portfolio #25) persistently lagged other similar funds in rankings over several quartiles. There were no obvious issues; asset allocation choices were sensible and there were no overt sources of portfolio implementation shortfalls.

However, the analysis of the graph model for this portfolio's baseline workflows showed hot-spots centered on information flows specific to associating data regarding the investment strategy proper (risk parity) as well as its subsidiary themes (e.g., stable BB rating + cheap valuation) to the plethora of transaction, position, and valuation records that were being managed. This observation was corroborated when analysis of other portfolios run by the same manager (Exhibit 1: portfolio #28 and portfolio #36) pointed to a similar cause for the degradation in their DOT metrics.

By way of example, Exhibit 5 shows the critical segment of the baseline (unoptimized) data schema that underpins the prime broker's trading system. This schema segment materially conditions the content of the electronic records that move onward to other functional groups (adjacent nodes in the graph model, within and outside of the firm). The corresponding data schemas in the manager's downstream systems (e.g., portfolio management, risk management, settlements) were not well served by the ability of the prime broker's systems to satisfy the information needs regarding trades and the associated portfolio's strategy. There were varying degrees of declining information fidelity as data transited across nodes and paths that permeated the entire workflow—most notably with block/basket trading allocations (DOT = 0.72), and the optimizers and impact on rebalancing (DOT = 0.58).

Remediation of the data schema issue required modifications to the data management environments in trading, portfolio management, and settlements—for the prime broker and the manager. The revised schemas that were deployed offered better information concordance between sending and target nodes. The revised graph model forecast a 166 bp potential gain, with the actual reported gains coming to 180 bp. Of all the workflow remediation alternatives, node reinforcement yielded the best improvements in portfolio returns, with the observed maximum of 180 bp, versus 166 bp for edge reductions and 128 bp for node reduction (annualized, gross of fees).

In general, schema upgrades typically involved modifying both the source and the target nodes to maximize the fidelity of information transfer. This can be accomplished by removing or minimizing data disconnects—either information from the source that could not be digested by the target, or conversely, the source being deficient in terms of discrete data required by the target. For example, we saw recurring instances where significant information (e.g., the revised cashflows stemming from realized extension risk in RMBS deals) was broadcast in free-form “Comment” or “Special Instructions” fields rather than in more rigorous data structures.

Disjointed schemas, which are markers of poor information flows, often were associated with implementations that relied on an amalgam of disparate systems arising inadvertently from prior organizational mergers or multiple layers of vendor acquisitions. We saw these repeatedly across many of the systems supporting the workflows we analyzed.

Tuning schemas are largely a manual process today . . . still more art than science. However, a promising technology under development that could make harmonizing schemas both simpler and more rigorous stems from research in a branch of mathematics called Category Theory (Bakclawski et al. 1994; Johnson 2001; Piessens and Steegmans 1995; Schultz and Wisnesky 2017). In effect, Category Theory ensures that one can *mathematically* guarantee correctness of data transformations from source to target schema, as well as embed data provenance within those transformations. If Category Theory can scale up to cope with the schema complexity we see in graph models of investment workflows (i.e., tens of thousands of communicating nodes), then it can be applied to automate the substantial challenges implicit in node reinforcement initiatives based on optimizing data schemas.

Information Parity

One novel outcome conveyed by our latest case studies showed that the best approach to optimizing operational efficiency (characterized with low DOT values) was to ensure that all the paths in the network (or at a minimum, those associated with nodes of high importance or “centrality”) were roughly on par in terms of DOT, within a range of ±15%. Beyond this band, further dispersion in DOT values likely generated impedance mismatches in information throughput, resulting in efficient nodes overwhelming less efficient ones. Wide dispersion in DOT values (implying some nodes were dramatically more efficient than others) always were associated with significantly more reconciliation breaks.

As a result, *information parity* emerged as a recurring objective in the optimizations across many of the firm-specific graph models. Some nodes were, by design, not maximally optimized even though the effort would have been very manageable (e.g., by simply adopting a more current messaging standard to better service more novel security types, for example). In effect, information parity ensured that the overall performance of the firm’s operations could achieve a level of equilibrium suited to the volume, velocity, and variety of data coursing through the entire network. By maximizing the likelihood that all parties to a portfolio's implementation had the same information at the same time, the level of reconciliation required was also materially reduced.

Notably, attempts to optimize the network in terms of seeking the shortest path for information traversal for each security type (addressing the "traveling salesman" problem that is a staple in operations research) yielded only half the improvements in DOT and subsequent alpha recovery. In effect, efficiency and optimality should not be conflated. As a result, aiming for information parity, rather than economy, is the more important driver, and supports the intuition that the coherent flow of information across nodes is more critical to the overall operability of information dissemination across the firm and with its counterparties.

Limitations

Approximately 4% of the nodes or edges within each firm's graph network model had no observable data of sufficient quality for analytic purposes despite their critical nature, as expressed in the investment agreements and guidelines. Typically, these were due to processes that were predominantly manual, such as cash inflows communicated by the institutional investor or their consultant via electronic mail to the Relationship Managers at the investment firm, or conducted largely through low-touch/low-value relationships, such as large-cap US equity trades using agency brokers. Therefore, we could not evaluate their specific impacts to the overall graph network metrics for that particular strategy at the investment firm.

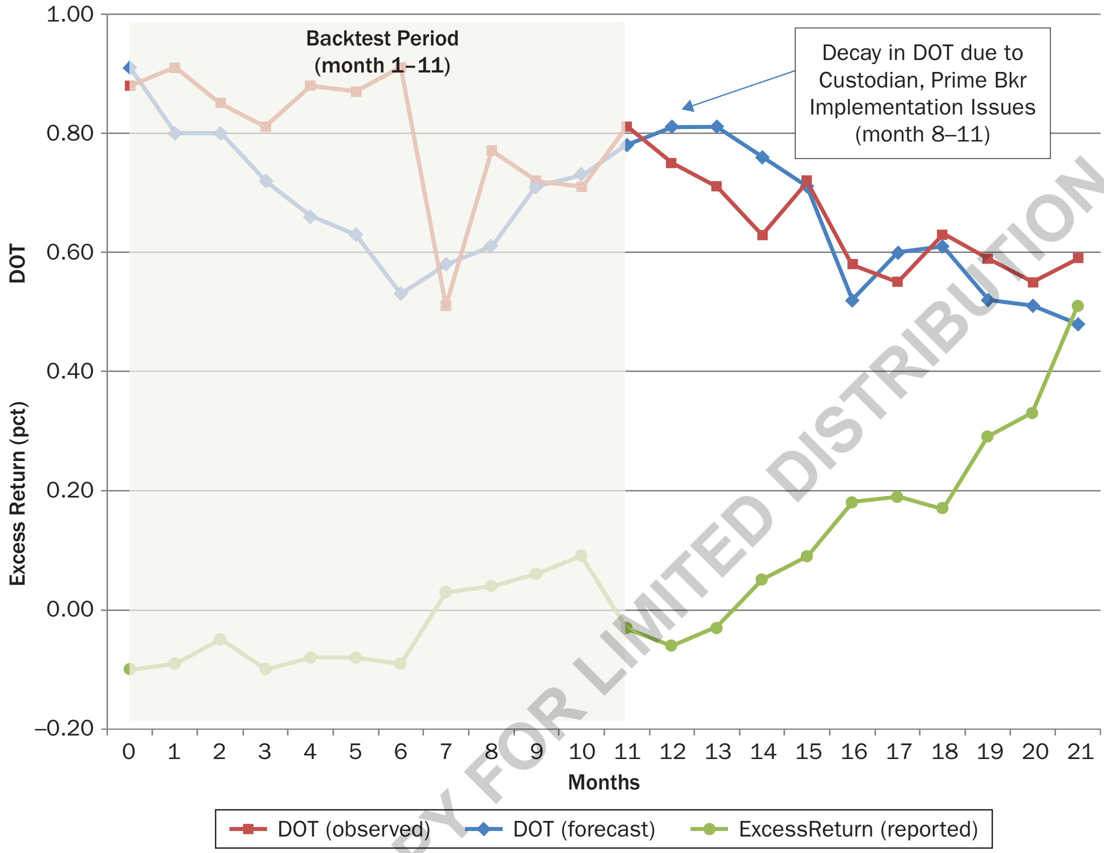
Our current study did not have the time to fully elucidate all the structural features of the various firm-specific graph networks. However, we took note of intriguing elements for future research such as:

- bellwether community structures (e.g., "bowtie"-shaped flows that reflect the centralized role of prime brokers in presenting price information)
- unique interactions across network layers (e.g., across the capital structure of an issuer and its influence in setting operational decisions for liquidity metrics)
- information diffusion processes (e.g., less central nodes exert material impacts in amplifying or attenuating "contagion," such as the propagation of erroneous data, or their measurable contribution to cumulative lags in data timeliness and distribution)
- senescent pathways (e.g., the impact of asset types with prolonged bouts of reduced liquidity and/or market activity on the quality of data workflows).

Approximately half of the graph network models for the firms we examined represented very complex implementations, with tens of thousands of nodes and edges across multiple layers. These translated into long simulation run times, from 40 minutes to 4.5 hours. We were able to ameliorate performance issues by deploying more system resources, such as securing processors and memory from commercial cloud-based services. However, throwing hardware at the problem is not a scalable response for the long term, and we continue to track a significant body of research that is looking to implement better graph model algorithms (e.g., McLaughlin and Bader 2018).

EXHIBIT 6

Long/Short Implementation: DOT Metric Analysis and Recovery in Alpha



NOTES: Months 1–11 was the backtest window exploring potential impacts of schema changes to the DOT metric and resulting portfolio performance. Implementation of data schema improvements into live workflow on Month 12 showed immediate benefits to the observed DOT metric and the corresponding excess returns reported. The decay in DOT in Months 8–12 was due to transient operational issues at the custodian and prime broker, which caused alpha to degrade. Resolution of the problems by Month 12 generated sustainable improvements to excess returns going forward.

DISCUSSION

This study demonstrates that minimizing the dissipation of alpha stemming from operational inefficiencies can resurrect a significant slice of the potential risk-adjusted returns in an investment strategy. This is especially valuable in the context of today's low-return environment, combined with investors' relentless push for lower fees. Optimizing investment operations workflows may not be as alluring as the top-line challenges related to asset selection and allocation—but the potential benefit of preserving 24-242 bps in a portfolio's return make for a compelling value proposition nonetheless.

In addition to re-affirming the relationship we observed previously between efficient investment operations and portfolio performance (Jovellanos 2011), we also describe novel applications of established tools from the graph modeling domain that institutional investors and their investment managers can bring to bear on:

1. establishing the optimal portfolio implementation workflow that can best express the alpha inherent in a strategy;

2. enabling a mechanism for continuous oversight of operational effectiveness by tracking sequential and cascading sources of portfolio implementation shortfalls; and
3. providing quantitative inputs (via the DOT metric) into their optimizers and risk models to enable tactical adjustments that can offset performance gaps.

Our latest observations strengthened what we had surmised previously (Jovellanos 2011)—that investment operations workflows are optimal when they are designed to explicitly service individual portfolio strategies (such as risk parity) rather than simply aligning along undifferentiated tasks or administrative functions (such as settlements). In effect, when it comes to establishing operational workflows within a firm, one size does not fit all strategies.

Operational Workflow and Investment Performance

The intuitive yet subtle nature of linkages between operational workflows and investment outcomes can be gleaned from the construct of the Information Ratio (IR) (Grinold 1989). Recall that IR represents excess return relative to tracking error, with the prevailing form of the equation:

$$IR = IC * sqrt(N) * TC$$

where IC is the Information Coefficient (measured as the correlation between the manager's expected excess returns and the subsequent excess returns of the assets in the portfolio), and N is referred to as "breadth" or the number of independent bets in the portfolio—though usage in practice typically falls to the number of tradeable instruments in the investment universe.

The key term for the purposes of this discussion is the Transfer Coefficient (TC), which reflects the efficiency with which portfolio managers translate their insights into portfolio weights (Clarke et al. 2002). TC equals one in the absence of constraints (such as restrictions imposed by pension funds per the investment agreements and guidelines), while adding constraints causes TC to decline and consequently impair IR.

An informal sampling of published IR values from 19 portfolios in both our current and 2011 studies suggest that DOT and TC are inversely correlated ($r^2 = -0.73$). The higher the DOT value (and the less efficient information flows are), the lower TC becomes, indicating higher obstacles for managers to express their insights into realized portfolio positions.

The linkage between TC and DOT was made clear in the context of relaxing the short sale constraint, whereby an investor can typically increase risk and returns without reducing TC and compromising IR in turn. In addition to fees from participation in lending activity, signals from the broader lending market provide material inputs into larger allocation decisions. This investment perspective motivated the implementation of a long/short portfolio (Exhibit 1: portfolio #37) by an institutional investor in our study.[7] Portfolio performance improved (51 bps) when the analysis of the DOT metric in the existing operational workflow highlighted shortfalls in the "locate" capabilities for specific securities across different market venues. The institutional investor recognized this operational deficiency and tasked a different prime broker to provide borrowing and lending services. Exhibit 6 shows the trajectory of forecast and observed DOT as well as the improvements in excess return.

[7] See Molk and Partnoy (2019) for a broader discussion on the merits of shorting for plan sponsors.

Looking Ahead

After 20 years of observation across 210 portfolios from 32 managers, we have been afforded a unique opportunity to study in detail the portfolio implementation process and the measurable impact of workflows on the expression of alpha from an investment strategy. Though it was inherently difficult, we were able to approximate "experiments" in the real-time, real-world environment of investment management (Lopez de Prado and Lewis 2018).

We remain hopeful that serendipity will continue to work in our favor and enable us to observe more portfolios over the long term. In addition to deepening the pool of portfolios studied, we wish to continue exploring how durable the benefits of workflow revisions are and their contribution to sustaining the long-term performance of investment strategies.

REFERENCES

Backlawski, K., D. Simovici, and W. White. 1994. "A Categorical Approach to Database Semantics." *Mathematical Structures in Computer Science* 4: 147–183.

Blackwell, J. 2014. "IBOR: Many Answers to Many Problems?" *CutterAdvantage* 97 (September). https://www.cutterassociates.com/cutter-advantedge/issue-2014-9.cfm.

Clarke, R. G., H. De Silva, and S. Thorley. 2002. "Portfolio Constraints and the Fundamental Law of Active Management." *Financial Analysts Journal* 58 (5): 48–66.

Grinold, R. C. 1989. "The Fundamental Law of Active Management." *The Journal of Portfolio Management* 15 (3): 30–37.

Hegeman, T., and A. Iosup. 2018. *Survey of Graph Analysis Applications*. https://arxiv.org/abs/1807.00382.

Johnson, M., and C. N. G. Dampney. 2001. "On Category Theory as a (meta) Ontology for Information Systems Research." *Proceedings of the International Conference on Formal Ontology in Information Systems* 59–69.

Jovellanos, C. 2011. "The Impact of Investment Operations on Portfolio Performance." *The Journal of Investing* 20 (3): 40–52.

Lopez de Prado, M., and M. J. Lewis. 2018. "Detection of False Investment Strategies Using Unsupervised Learning Methods." https://ssrn.com/abstract=3167017.

McLaughlin, A., and D. A. Bader. 2018. "Accelerating GPU Betweeness Centrality." *Communications of the ACM* 61 (8): 83–92.

Molk, P., and F. Partnoy. 2019. "Institutional Investors as Short Sellers?" *Boston University Law Review* 99: 837–871.

Perold, A. 1988. "The Implementation Shortfall: Paper versus Reality." *The Journal of Portfolio Management* 14 (1): 4–9.

Piessens, F., and E. Steegmans. 1995. "Categorical Data Specifications." *Theory and Application Categories* 1 (8): 156–173.

Robertson, A. 2019. "Passive in Name Only: Delegated Management and 'Index' Investing." *Yale Journal on Regulation* 36 (2). https://digitalcommons.law.yale.edu/yjreg/vol36/iss2/6.

Schultz, P., and R. Wisnesky. 2017. "Algebraic Data Integration." *Journal of Functional Programming* 27. https://doi.org/10.1017/S0956796817000168.