

---

**JOIM CONFERENCE SERIES  
SEPTEMBER 21–23, 2014  
DOUBLETREE BY HILTON SONOMA  
WINE COUNTRY HOTEL**

***Data Science, Institutional & Personal Investing***

**PRACTITIONER'S SUMMARIES**

---



**Arthur B. Laffer, Laffer Associates**

*The Wealth of States*

This talk, in tandem with Dr. Laffer's newest book, *An Inquiry into the Nature and Causes of the Wealth of States*, coauthored with Travis Brown, Rex Sinquefeld and Steve Moore, explains why eliminating or lowering tax burdens at the state level leads to economic growth and wealth creation. As argument for tax reform, Dr. Laffer will show that states of all sizes can benefit enormously with the right policies. His detailed exposition evaluates the impact state and local government policies have on a state's relative performance and economic overall growth.

**Vineer Bhansali, PIMCO**

Speaker

*By the Numbers: 10 Things my Hobbies have Taught Me About Investing*

Discussant: Jing Zhang, Moody's Analytics

I discuss ten common themes between non-investment related activities that shed practical and useful light on investing. While readers might not be familiar with these particular activities, I believe that combining analogies from any accumulated skill in intrinsically rewarding activities (also known as hobbies), with a disciplined analytical approach yields significant benefits. The themes (including a bonus eleventh) are:

- Focus on Structure
- Use Data and Bayes rule
- Use Proper Units
- Identify repeated Type I and II errors
- Simplify
- Identify scaling and sharp transitions
- Simulate
- Make good tradeoffs
- Don't fight momentum
- Watch the environment
- Use commonsense: If its too good to be true it probably is

**Vasant Dhar, New York University**

Speaker

*Data Science and Prediction*

Discussant: Peter Lee, AlphaSimplex Group, LLC

Data science is the study of the generalizable extraction of knowledge from data. A common epistemic requirement in assessing whether new knowledge is actionable for decision making is its predictive power, not just its ability to explain the past. The heterogeneity and scale of data and diversity of analytical methods require data scientists have an integrated skill set, as well as a deep understanding of the craft of problem formulation and the science required to engineer effective solutions. I discuss the potential and limitations of Data Science in building predictive models in Finance.

**Mark Kritzman, Windham Capital Management, LLC**

Speaker

*The Components of Private Equity Performance: Implications for Portfolio Choice*

Discussant: Jim Quinn, CalPERS

We use a proprietary database of private equity returns to measure the excess return of private equity relative to public equity covering a period of approximately 17 years. We then partition this excess return into two components: an asset class alpha and an illiquidity premium.

Our empirical analysis offers persuasive evidence that the sector weights of private equity funds do indeed predict the subsequent performance of public equity sectors within both large and small capitalization universes. This means that investors can capture this asset class alpha using sector ETFs to match the sector weights of private equity funds without incurring the encumbrance

of illiquidity. We refer to this strategy as liquid private equity.

We also show how this decomposition of the private equity excess return into an asset class alpha and an illiquidity premium affects the optimal composition of a portfolio. Private equity loses appeal relative to liquid private equity because liquid private equity delivers the asset class alpha without subjecting the investor to illiquidity. And public equity becomes less attractive relative to liquid private equity because it does not offer an asset class alpha. Finally, private equity becomes less attractive relative to all liquid asset classes, because it is now seen to offer a smaller premium to compensate for its illiquidity.

Co-authored with William Kinlaw and Jason Mao

**David Leinweber, Berkeley National Laboratory**

Speakers

*Event Driven Trading and the “New News”*

Discussant: Ronald Kahn, BlackRock

This talk is an update of “**Event Driven Trading and the ‘New News’** (Leinweber & Sisk, *Journal of Portfolio Management*, Vol. 38, No. 1, 2011). There are two information revolutions underway in trading and investing. Recent headlines focus on structured quantitative market information at ever higher frequencies. The other technology revolution is driven by qualitative, textual and relationship information. The IBM computer Watson’s overwhelming Jeopardy victory demonstrated how good machines can get at this. News analysis is a focus of language technology in finance.

A key finding regarding “basic news”, the kind that appears as a story in a newspaper or on a website, is that there is a great deal of truth to the old adage, “buy on the rumor, sell on the news”. A earlier study (Tetlock *et al.*, *J. Finance* 63, 2008)

of 20 years of news, from 1984–2004, found that textual “sentiment analysis” of news stories did in fact predict with high accuracy that upward stock price movements were associated with positive news, and downward moves with negative news. However, well over 90% of these news related price moves occur *before* the publication of the news. Almost always, someone knows the news before it appears, and that knowledge is impounded in the price.

The 2001 paper referenced in the title found that the same observations were evident in “newer news”, well into the modern web era, from 2003–2008. However, more complex strategies beyond “buy on the good news, sell on the bad” did produce exploitable alpha, up to 10%/year for a period of two years out of the sample. Given the ever-shorter times that technology brings to trading strategies, there is no doubt a race to the bottom in latency, and the remaining alpha in even complex reactive news strategies is reduced, and shared among the fastest participants.

Newer efforts at automated qualitative analysis, go beyond traditional news as a core source. Some use “pre-news” such as legal and regulatory actions that have a slower migration into mainstream news. Others rely on social media, such as twitter and search trends. These tend to have a narrow breadth, focusing on stocks of particular interest to the public.

Perhaps the most promising path to exploitable information is from the revolution in imaging data. The new generation of small earth-looking satellites provides a standardized global set of “big data” that can be analyzed using a mix of human and machine methods. This can illuminate changes in firms’ logistics and supply chains; information on commodities in the form of crop yields, mining activity from tailings and traffic, and global shipping activity. This information is just starting to be available. They used

standardized methods and are not subject to government “editorializing” or tilts as is the case today. As baselines are established they will be used more extensively in investment management.

**Tim Loughran, University of Notre Dame**  
Speaker

*Measuring Readability in Financial Disclosures*

Discussant: Tal Sansani, American Century Investments

Defining and measuring readability in the context of financial disclosures becomes important with the increasing use of textual analysis and the SEC’s plain English initiative. We propose defining readability as the effective communication of valuation-relevant information. The Fog Index—the most commonly applied readability measure—is shown to be poorly specified in financial applications.

The Fog Index indicates that an increase in the number of complex words (more than two syllables) decreases readability, with this factor accounting for half of the measure’s inputs. Business text, however, commonly contains multi-syllable words used to describe operations. Words like *corporation*, *company*, *agreement*, *management*, and *operations* are predominant complex words occurring in annual reports, yet are presumably easy for investors to comprehend.

As a measure of readability in financial disclosures, we recommend that researchers use the file size of the “complete submission text file” available on the SEC’s EDGAR website. In sum, a simple measure like annual report file size appears to better capture how effectively managers communicate valuation-relevant information to investors as measured by subsequent stock return volatility, earnings surprises, and

analyst dispersion than a traditional readability measure like the Fog Index.

co-authored with Bill McDonald

Journal of Finance, Volume 69, Issue 4 (Aug 2014), 1643–1671.

**Ananth Madhavan, BlackRock, Inc.**

Speaker

*Understanding ETF Price Dynamics*

Discussant: Kenneth Blay, 1<sup>st</sup> Global

Exchange traded funds (ETFs) have grown substantially in size, diversity and market significance in recent years leading to increased attention by investors, regulators and academics. While the benefits of ETFs in terms of offering a large diversity of exposures at low cost are now well understood, concerns about the pricing and trading of these investment vehicles have also been voiced. Common themes include the transmission of liquidity shocks via arbitrage, excess volatility, and economically significant premiums/discounts, particularly in times of market stress. In particular, premiums and discounts in ETFs are a persistent source of questions because there is no such analogue for an open-ended mutual fund. Specifically, why do ETFs trade at premiums or discounts? How should investors incorporate these into their investment decisions? How do pricing errors get corrected and how quickly does this take place?

Our research develops a model of ETF price dynamics based on the arbitrage mechanism unique to ETFs. The model yields a natural metric for the speed with which arbitragers act to correct deviations between the ETF's price and intrinsic value. Using a state-space representation, we show how to decompose the ETF's price premium to its Net Asset Value (NAV) into price discovery and transitory liquidity components. We use this decomposition to generate a measure

of price discovery. We then estimate the model individually for all 947 US-domiciled ETFs in the period 2005–2014 with a full year of trading history. We find that arbitrage speed varies widely across funds and exposures, and is systematically related to cross-sectional measures of liquidity. We provide an illustration of a bond ETF during the financial crisis of 2008 to highlight how apparently dramatic discounts really reflected price discovery when the underlying basket was illiquid in the extreme. The decomposition of the premium into its components is important because investors may avoid buying at a premium or selling at a discount when in effect the price of the ETF has moved to capture changes in unobserved value. It also highlights the point that seeming tracking error of an ETF is not “mispricing” but often, especially in times of market stress, a reflection of price discovery by the more liquid instrument.

**Jose Menchero, MSCI**

*Efficiently Combining Multiple Sources of Alpha*  
Speaker

Discussant: Margaret Stumpp, Quantitative Management Associates

The starting point in active management is to identify the sources of alpha for the portfolio. However, even a high quality alpha signal may result in mediocre risk-adjusted performance if not properly implemented. We begin our study by investigating and comparing various portfolio construction techniques for capturing a single source of alpha. The most efficient of these portfolios is given by the minimum-volatility factor portfolio. This portfolio is formed by mean-variance optimization and has the highest expected return per unit of risk.

While the solution of the minimum-volatility factor portfolio has been known for some time, many practitioners regard the resulting portfolios with

suspicion. As a result, they often impose investment constraints on the solution to obtain more intuitive factor exposures. We argue that this practice may increase the risk of the portfolio, without boosting expected returns. One aim of our paper is to develop the intuition behind these exposures and demonstrate that such exposures can be quite effective at reducing portfolio volatility.

Finally, we turn our attention to the problem of efficiently combining multiple sources of alpha. We derive the intuitive result that the optimal portfolio for a composite alpha signal is simply a weighted combination of the minimum-volatility factor portfolios for each alpha signal taken separately. We show how to compute the appropriate weights and further demonstrate our technique with a detailed example.

Forthcoming in the Journal Of Investment Management (JOIM)

**Nancy Wallace, University of California, Berkeley**

Speaker

*Endogenous Financial Norms in Financial Networks*

Discussant: Roger Stein, Massachusetts Institute of Technology

We develop a theoretical model of a network of intermediaries, which we apply to the U.S. mortgage supply chain. In our model, heterogeneous financial norms and systemic

vulnerabilities arise endogenously. Intuitively, the optimal behavior of each intermediary, in terms of its attitude toward risk, the quality of the projects that it undertakes, and the intermediaries it chooses to interact with, is influenced by the behavior of its prospective counterparties. These network effects, together with intrinsic quality differences between intermediaries, jointly determine financial health, quality, and systemic vulnerability, at the aggregate level as well as for individual intermediaries. Our model therefore allows us to evaluate the relative importance of network effects for intermediary behavior. It also makes predictions about the conditions under which a network is systemically vulnerable to shocks and about how such systemic vulnerabilities can be detected empirically. We apply our model to the mortgage-origination and securitization network of financial intermediaries, using a large data set of more than 12 million mortgages originated and securitized through the private-label supply chain from 2000–2007. We then track the ex-post foreclosure performance of each loan in the network and compare the evolution of credit risk by vintage with the model's predictions. We find that credit risk evolves in a concentrated manner among highly linked nodes, defined by the geography of the network and the interactions between originator and counterparty over time. This suggests that network effects are indeed of vital importance for understanding the U.S. mortgage supply chain.