New Economic Models for Labour Dynamics of the Green Transition

Transfer of Status Report to the School of Geography and the Environment

Owen Haaga

July 13th, 2024 (Revised)

Contents

Outline of Research Plans

Models for the Transition

The political-economic challenges of the green transition will require timely, credible, and detailed modeling of labor and product markets, and of the macroeconomic feedback effects of foreseeable disruptions to them. This thesis will demonstrate the use of emerging modeling techniques to deliver policy-relevant forecasts and counterfactual analysis of the complex interacting dynamics involved.

Pandemics, disruptions in commodity markets and supply chains, and increasing regional and individual disparities in OECD countries all require economic policymakers to act on knowledge which conventional policy and forecasting models fail to deliver [see, for example, Farmer, et. al., 2015, and Stiglitz, 2018]. Anthropogenic climate change is only heightening these demands [Stiglitz, Stern, and Taylor, 2022]. Fortunately, advances in computation, machine learning, and modeling of complex and spatially embedded economic systems promise to enable the construction of a new generation of forecasting and policy models.

Due to the complex interrelationships of highly disparate environmental and social systems, it is imperative that modeling approaches enable interdisciplinary collaboration and reciprocal interaction with stakeholders. Given the contentious political economy of the policies involved, a successful model will also provide relevant levels of geographic detail, and be clearly driven by authoritative and publicly available data.

Existing Mainstream Efforts

Existing macroeconomic models in the dynamic stochastic general equilibrium (DSGE) tradition fall short on several of these dimensions. They are heavily assumption-driven, presented in terms that are opaque to non-specialists, and fundamentally very difficult to combine in a meaningful way with models from other disciplines. Crucially, the assumption of period-by-period equilibrium in the labor market renders these models inadequate for modeling adjustment processes to labor market change, or persistent unemployment in any form [Stiglitz, 2018].

Because of these shortcomings, a largely separate literature has developed within mainstream economics literature to model labor markets. Although "search and matching" models retain equilibrium assumptions in other markets, they impose a "black box" matching function in the labor market, which directly specifies how job seekers match (or fail to match) to vacancies, typically based on the ratio between the two [Mortensen, 2011]. Economists working in this tradition have increasingly emphasized a "flow-based" approach to modeling the labor market [Ebrahimy and Shimer, 2010], and statistical agencies in the United States such as the Bureau of Labor Statistics and the Census Bureau have responded to this demand by releasing increasingly detailed official estimates of labor market flows [see, for example, https://www.bls.gov/jlt/].

These estimates, in turn, have not only begun to inform the public commentary around the state of the labor market, but also fed a further turn towards empirical modeling in the literature. Often with limited or no appeal to a "search and matching" theoretical model, researchers have generated meaningful insights from observing flows of individuals between different labor market states (employed, unemployed, out of the labor force) and using the structure of the network of flows to draw out implications for the time path of variables of interest [for example Elsby et. al., 2019].

Convergent Developments in Systems Dynamics and Statistics

Although the two literatures do not yet seem to intersect, this style of modeling which has recently borne fruit in the investigation of labor market flows has a long history both in heterodox economics and much more widely, under the name of "systems dynamics" [Radzicki, 2009]. While the models used so far in the labor flows literature have remained relatively simple, academics and practitioners working in the systems dynamics tradition have built tooling and practices that scale to much larger and more detailed models. Importantly, this tradition has placed a great deal of emphasis on participatory modeling with interdisciplinary domain specialists and stakeholders [Barbrook-Johnson and Penn, 2022]. Unsurprisingly, these methods have achieved widespread use in situations related to combined social and environmental systems [Radosavljevic et. al., 2023].

Unfortunately, the systems dynamics community has been somewhat divided on the desirability of treating their models as statistical objects to be estimated from data [Sterman, 2018]. These models have also traditionally been built up by hand, even though they can become unwieldy as more states and concepts are added. New developments in statistical programming, inspired by category theory, promise to overcome these hurdles. In particular, the ability to compose small models into larger models promises to make building complex models much more tractable [Baez et. al., 2022].

By building models using categorical methods the Julia programming language, we can seamlessly apply state of the art solution and estimation methods for our models, by interpreting them as systems of (stochastic) differential equations [Meadows, Li and Osgood, 2023]. Many recently developed methods suggest themselves, such as symbolic regression and the various neural network methods grouped under the term "Scientific Machine Learning" [see https://sciml.ai/].

Empirically grounded systems dynamics modeling of labour markets should generate reliable flow forecasts as well as measures of normal variability for quantifying the ability of these markets to absorb sector-level shifts in labour demand. The output from this modeling excercise should also provide a rich set of targets for calibrating individual and household-level models connecting realistic behavioural rules (and microdata) to aggregate dynamics. This fine-grained understanding of labor market flows will provide a crucial input to the process of Agent Based Modeling to tease out local macroeconomic and household welfare implications of large changes in local labour market demand, and to provide quantitative guidance for high-stakes industrial policy in the context of complex interacting dynamics.

Fundamental Research Question

What can historical patterns of labour market adjustment tell us about the dynamics of unemployment in the Green Transition? Firstly, what underlying dynamics account for the regularity in the speed of labour market recoveries in recent data (with the partial exception of the pandemic)? Secondly, how well can the available disaggregated data allow us to forecast high-frequency dynamics of unemployment in the near term, and place bounds on the realistic speed of adjustment to longer-term changes in labour demand which are foreseeable as part of the Green Transition?

The labour market modeling portion of this thesis seeks to interpret observed regularities in labour market dynamics in terms of an underlying stock-flow model, and to derive empirical estimates of the relationships in that model. As illustrated and outlined below, some of these observed regularities are of long standing, and others have been discovered much more recently. A fully estimated, flow-based model will make it possible to translate hypothesized changes in the regional and industrial pattern of labour market demand (i.e. Green Transition scenarios) into quantitative forecasts of labour market dynamics.

Specifically, I will build a systems dynamics model of flows between employment, unemployment, and out of the labour force using monthly flow data at the national level from 1990 onward, and iteratively add industry, occupation, and geographic detail to the model using other publicly available stock and flow data at different time resolutions and levels of detail. The software I will use should allow me to impose adding-up constraints and known model relationships with certainty, and to compose the larger model by "stratifying" smaller and more tractable models along different dimensions. The available data (more than 400 observations for each flow in the initial model, for example, with more data available for each new dimension of the iterative process) should be more than sufficient to estimate plausible linear relationships. It is a more open question whether the data will be sufficient to support more sophisticated model selection to identify non-linear relationships. This may well depend on the precise methods used, and will be investigated in the second (methods) paper.

Outline of Papers

The first paper will involve constructing a proof-of-concept model of flows into and out of labor market states, occupations, industries, and sub-national geographic units. A flexible, stock-flow based framework combined with machine learning techniques (linear and possibly non-linear estimation methods) should make it possible to exploit the inherent structure of available official statistics from the United States, and to investigate recently observed patterns in unemployment dynamics which suggest complex dynamic behaviour. Which industries, regions, and processes have generated the observed regularity of a constant, steadily decreasing unemployment rate during recent business cycle recoveries in the United States? What does this regularity tell us about the resilience of the labour market to sectoral shocks?

The second paper will further explore and develop the methods involved in estimating the model from the first paper. The principal comparison will be between estimation methods based on genetic algorithms (symbolic regression), neural networks ("SciML"), and path signatures. Outcomes of interest include both prediction of the fundamental transition rates and ability to scale to more detailed data sources. Possible extensions include "borrowing power" from estimation of transition models from microdata, as well as investigation of flow models incorporating fractional-order derivatives. Can interpretable models match (or even outpredict) "black box" methods on the datasets available for labour market analysis?

The third paper will employ (and further develop) an existing agent-based model which combines labor and product markets with a model of housing and firm finance. This model, the "Macro ABM" under development at INET, aims to generate realistic industry and employment dynamics, and is calibrated to firm size distributions and financials. Given plausible assumptions on diminishing marginal propensities to consume, as well as "home bias" in investment, this model should be sufficient to roughly quantify the effects of a hypothesized link between income inequality and asset price inflation. This, in turn, would provide rough bounds on a pool of funds which could be available for redirection towards green investments with no first-order loss of economic efficiency (and possible large gains). How much of the green transition can be financed without compromising other forms of productive investment?

Preliminary Review of Literature

It goes almost without saying that labor markets will be a first-order concern in the political economy of the green transition. Although some indicators would suggest that, at least within the United States, opportunities in new energy sectors may align well (at least spatially) with the areas which will need to replace jobs in fossil fuel extraction [Curtis and Marinescu, 2023], the spectre of the "China Shock" hangs over any discussion of large scale economic transition. The failures of labor markets (or policy) to adjust to the deindustrialization of the American Midwest [Autor, Dorn, and Hanson 2021], has had echoes and analogues in many developed economies. In the case of the United Kingdom, the decline of the coal extraction industry has had similarly stark consequences for the regions most effected, with troubling implications for that aspect of the green transition elsewhere [Aragon et. al., 2022]. Together, these traumatic experiences of large-scale structural change in the recent economic history of a number of major economies provide a warning, and hopefully lessons, for the path ahead.

Unfortunately, our ability to make use of this experience will depend on a sound understanding of labor market dynamics, which the consensus of the economics profession has conspicuously failed to provide over the last few years. Even before the pandemic, policymakers had lost faith in the baseline macroeconomic model as a practical guide to policy [Board of Governors of the Federal Reserve System, 2018], and the dramatic recovery of the labor market during the pandemic defied most (but not all [Pichler et. al., 2022]) forecasts. The experience with inflation, and in particular the comovement of inflation and unemployment, in the subsequent years has further underlined the inadequacy of our current understanding of the limits of labor market adjustment.

The most promising basis on which to reconstruct our understanding of the labor market is the so-called "flow-based approach", named by [Blanchard and Diamond, 1992], but largely associated in recent years with the "search and matching" modeling tradition inaugurated by [Mortensen and Pissarides, 1994]. Following [McCall, 1970], this tradition relaxes the (glaringly counterfactual) assumption of period-by-period labor-market clearing, at the cost of introducing an ad-hoc matching function which determines the extent of the disequilibrium.

Although representing an incomplete step towards fully general disequilibrium modeling, this literature has not only engaged seriously with existing empirics, but also spurred the development of high-quality official data sources on flows between employment, unemployment, and outside the labor force. Statistics derived from the Job Openings and Labor Turnover program at the Bureau of Labor Statistics, in particular, provide high-frequency estimates of the key quantities of interest. Recently released experimental estimates at the State level have taken a further step towards providing a meaningful model of local labor markets in the official statistics [available at https://www.bls.gov/jlt/jlt_statedata.htm].

Economists pursuing a flow-based approach to labor market analysis have shed light on several phenomena which would have been inaccessible to analyses which disregarded the stocks and flow structure of the quantities involved. Examples include the importance of labor market "churn" and flows into employment from outside the labor force [Abraham, Haltiwanger, and Rendell 2020], and the "participation cycle" through which business-cycle fluctuations in employment vs. unemployment drive longer-term deviations from the trend in labor force participation vs. nonparticipation [Hobijn and Sahin, 2021].

One particularly promising line of research suggests that the dynamics of labor market recoveries can be largely seen as a steady (.1 log percentage point) downward drift in the unemployment rate in between sharp jumps during recessions [Hall and Kudlyak, 2022]. In fact, this observed regularity has been cited to advance the hypothesis that the natural rate of unemployment itself moves in this fashion, and that the fundamental limits on job growth are therefore best conceptualized in flows rather than levels [Hall and Kudlyak 2023].

To further investigate these and other regularities in the labor-flows data, economists have largely used estimates of individual heterogeneity from publicly available microdata, using traditional linear methods as well as some newer empirical clustering techniques. [Ahn and Hamilton, 2020, Hall and Kudlyak, 2022b, Ahn, Hobijn, and Sahin, 2023] These techniques, while allowing for the incorporation of valuable detail, essentially throw out the information contained in the accounting identities of the stock-flow framework, and ignore the demand-side of the labor market.

Although so-far unacknowledged in the job-flows literature, these developments have largely trended in the direction of the cluster of modeling techniques which have traditionally gone by the name of "Systems Dynamics" [Barbrook-Johnson and Penn, 2022, and many historic references such as Forrester, 2007 and Meadows, et. al., 1972]. Practitioners have developed techniques for conceptualizing, building, and validating stock and flow-based models which should prove extremely valuable in analyzing the full range of available flow-based macro labor market data. They have also, in many cases, demonstrated the complementary use of these methods with individual-level "agent-based" modeling techniques, which would seem to be the next logical step in the direction of exploring individual heterogeneity. One explanation for the failure of these traditions to intersect so far, apart from the usual tyranny of disciplinary boundaries, is the fact that Systems Dynamics practitioners have not always maintained a commitment to quantitative empirics, limiting contact with practitioners of data-driven social science from other backgrounds. Fortunately, a significant group within this community are open to (and prioritize) constantly incorporating new techniques to bring Systems Dynamics models to data [Sterman, 2018].

While various non-linear dynamic estimation methods and publicly available JOLTS data should allow us to estimate a richly detailed (and geographically specific) model of overall labor-market dynamics, the investigation of important aspects of demographic heterogeneity will require a complementary approach at the individual level. Instead of simply pulling estimates at the individual level from public use microdata, however, the stock-flow model can be used to calibrate a full agent-based model from that same microdata, which generates realistic detailed macro dynamics. Existing Macro and Labour models maintained by the INET Complexity Economics Group could be extended for this purpose, or else techniques under development by the Topos Institute could be used to rapidly develop a new model. Once developed, this model could then provide a link between labor market trends (such as increased income inequality) and household balance sheets, particularly housing markets. The hypothesized connection between the two could be of first-order importance in the redirection of surplus loanable funds to productive public investments rather than supporting unstable asset-price inflation [Mian, Straub, and Sufi 2020].

Proposed Contributions to the Literature

Despite the ubiquity of the Beveridge Curve (Figure 1) in practical economic policy analysis, significant work has only recently gone into taking it seriously as a reduced form model of the dynamic relationship between flows into and out of different labour market states [Barlevy et. al., 2024]. While papers such as [Hobijn and Sahin, 2021] have examined these relationships at the national level, and the same authors have begun to use public use microdata from the national labour force survey (CPS) to observe clusters of individual trajectories through this process [Ahn and Hamilton, 2020, Hall and Kudlyak, 2022b, Ahn, Hobijn, and Sahin, 2023], I am not aware of any work that exploits the stock-flow structure to build a coherent model of labour market flows using sub-national data (such as state-level and industry-level JOLTS data, and the partial demographic data available on the unemployment claims data series).

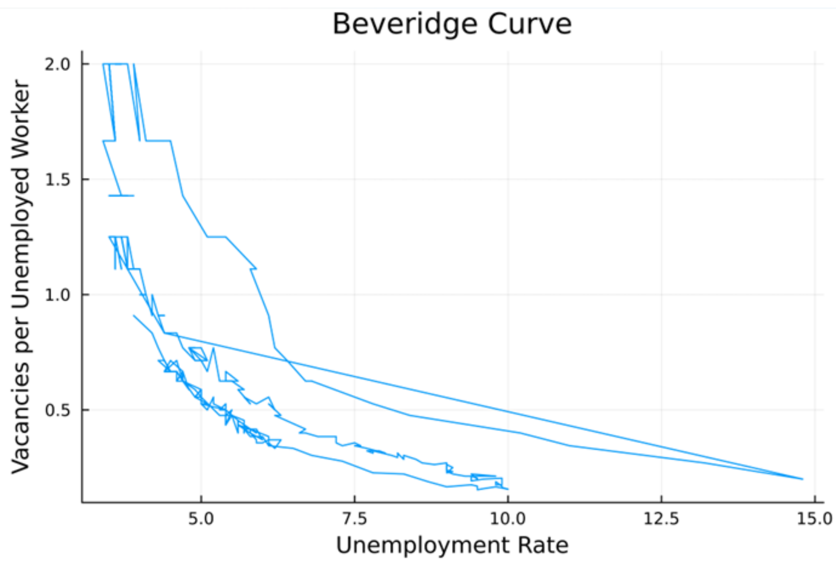

Figure 1. The "Beveridge Curve" shows a predictable relationship between labour market demand and unemployment, interrupted by shifts during recessions (author's calculations based on Bureau of Labor Statistics data from the Job Openings and Labor Turnover Survey).

In addition to exploring the underlying dynamics of the traditional Beveridge Curve relationship, this model would also shed light on the phenomenon (shown in Figure 2) recently brought to light (but not yet explained) by Hall and Kudlyak [2022], in which the post-war recoveries in the U.S. labour market took place at a strikingly constant rate (in log percentage points). Not only would a coherent view of the underlying flows structure which leads to this observed pattern be useful explaining it's regularity, it would also provide a useful way to investigate the contrast between these highly regular recoveries and the "steeper" recovery (followed by a recessionless rebound) of unemployment following the pandemic recession.

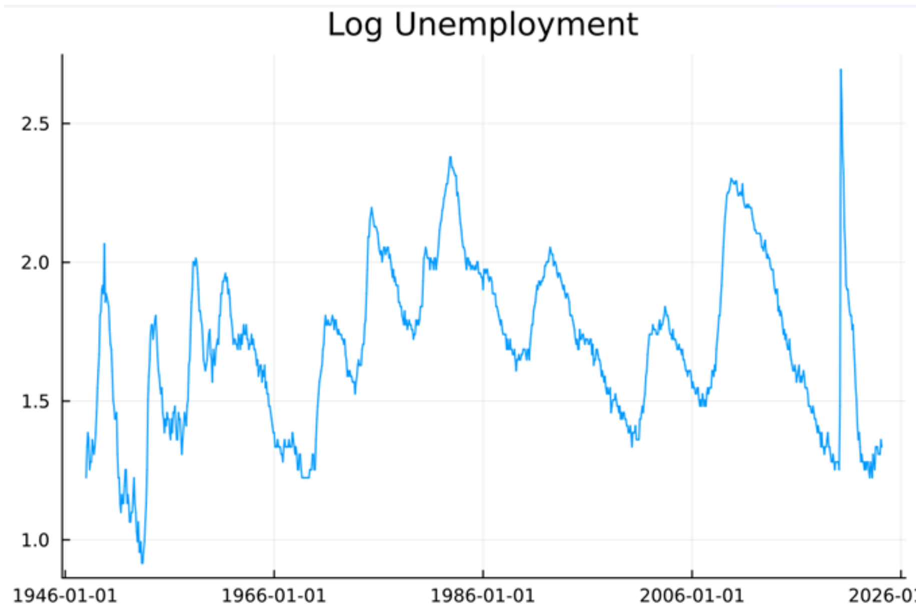

Figure 2. Hall and Kudlyak find that log percent unemployment drops linearly by about .1 per month in between upward spikes during recessions. The pandemic recession and recovery appear to be an exception (author's calculations based on Bureau of Labor Statistics data from the Current Population Survey).

Preliminary Review of Methods

System Dynamics

System Dynamics, and the "systems thinking" movement more generally, began with the application of then-novel methods from operations research and control theory to the analysis of social and natural systems in the early days of digital computing. Over time, the system dynamics community became associated very particularly with a style of model in which a set of real-valued state variables are conceived of as stocks, and interlinked by flows in between them, which vary subject to explicitly identified influences from other parts of the system. These influences may be specified in a quantitatively exact fashion, or (perhaps more commonly) they may be specified only up to the direction (sign) of the effect [Barbrook-Johnson and Penn, 2022].

For many practitioners, the System Dynamics tradition has come to be associated, in particular, with the qualitative analysis of models specified up to the sign of the influences involved. Identifying positive and negative feedback loops in the system, and deriving qualitative conclusions from their presence, has come to be a core discipline of systems thinking, sometimes referred to as “Causal Loop Analysis”. System Dynamics builds on this by providing tools for the analyst to assume functional forms for the influences and solve out the model as a system of equations. This allows for a qualitative exploration of the dynamics implied by the model, and is widely seen as the core of the discipline [Barbrook-Johnson and Penn, 2022].

One school of thought within System Dynamics encourages an even more quantitative approach, freely adopting the latest techniques from statistics (or “data science”) and new computational methods to derive more of the modeling process from data. These practitioners maintain that while modelers should never disregard qualitative knowledge about the system they are modeling, recent and rapid advances in quantitative methods present opportunities for System Dynamics practitioners which are fully in keeping with their principles and goals [Sterman 2018]. This is the approach I intend to take in this thesis, incorporating novel statistical methods and programming techniques in a largely empirical modeling process.

Despite their similarly quantitative approach, and a significant amount of overlap in the real world phenomena which they seek to explain, the System Dynamics tradition has had very little impact on the literature or methods of mainstream economics [Radzicki, 2009]. System Dynamics methods have been applied to labor market research (using calibration rather than the sort of direct estimation I propose here) [Jo, Kim, and Lee, 2023], and there are some similarities between the network-based ABM methods applied in [del Rio-Chanona et. al., 2021] and a system dynamics view of the labour market. System Dynamics, as such, however, is not mentioned at all in the economics literature on labor market flows referenced above, despite the obvious similarities in the modeling approach. While a full explanation would be beyond the scope of this document (and probably even of the thesis described herein), several partial explanations suggest themselves for the orthogonal development of these modeling traditions.

To begin with, the emphasis placed on stakeholder participation and incorporating qualitative knowledge ran counter to the trend of developments in economics during those years. In particular, the explicit rejection of methodological individualism, let alone the rational agent framework, put the System Dynamics community starkly at odds with the increasingly dominant paradigm in economics. Additionally, it must be said that the high-profile failure of the forecasts in the “Limits to Growth” report, although arguably ahead of their time in the effort to integrate environmental, social, and economic constraints, tied the System Dynamics tradition to a strain of behaviourally naive neo-Malthusian thinking in the 1970s (such as the infamous “Population Bomb”) which damaged its credibility in the eyes of many economists.

Although recent modeling excercises following up on the World3 model and its successors have pointed out that later versions of the model can give reasonable quantitative forecasts [Herrington, 2021], the original modeling excercise presented a fundamentally qualitative warning about the dangers of what it termed “resource depletion”. Economists at the time, such as [Solow 1973], were deeply skeptical of this argument, ironically because of an awareness of behavioural feedbacks in the system which would ultimately deliver an outcome other than that which could be foreseen by linearly extrapolating then-current trends. 50 years on, however, more abstract forms of resource constraint, such as biodiversity loss and the thermal properties of the atmosphere, threaten to constrain or collapse future levels of wellbeing.

Ultimately, the influence of System Dynamics in economics has remained almost entirely confined to the self-consciously heterodox tradition of “stock-flow consistent” modeling influenced by Wynne Godley and the Cambridge Department of Applied Economics [Radzicki, 2009]. The current generation of stock-flow consistent modelers, for instance those associated with the “Modern Monetary Theory” movement, have continued to position themselves as competitors to mainstream economists in the policy space, and do not seem to have made great strides in the directions advocated by the quantitative wing of the System Dynamics community. Ideally, by choosing to move in that direction, and addressing subject matter which even the mainstream has begun to address in a disequilibrium framework, this project proposed here would stand a better chance of meeting the orthodox economics literature half-way in a productive fashion.

Data

Partly in response to the growth of interest in the flow aspect of labor markets, the United States statistical system releases a wealth of publicly available data on stocks and flows in the labor force. These include the high-frequency headline releases from the Bureau of Labor Statistics, such as the series included in the Employment Situation Report, as well as the Job Openings and Labor Turnover Survey. BLS also releases lower-frequency (annual) data on employment and wages by occupation, and quarterly data on employment and wages by industry and firm size at low-level geography. The Census Bureau releases a great deal of information relevant to the labor market, including an anonymized sample of microdata from the primary labor force survey and a synthetic recreation of comprehensively linked data from numerous sources.

The closely-watched “Employment Situation” report from the Bureau of Labor Statistics, a monthly event that business journalists have taken to calling “Jobs Day”, draws on both the Current Population Survey (the venerable household labor force survey which has provided since its inception the official estimate of the unemployment rate), and an employer survey drawn from the sampling frame of the state unemployment insurance systems. These data series provide a wealth of information on stocks of employed and unemployed workers by industry, labor force, and demographic characteristics. While providing a timely and detailed picture of labor force aggregates, these statistics tend to be much stronger on marginal distributions than on cross-tabulations across multiple dimensions. Two weeks after each main release, a state-level employment report follows, further subject to this caveat for being tabulated at a finer level of geography. Famously, quality and statistical disclosure concerns prevent BLS from releasing even an unemployment rate estimate by both state and race.

The Job Openings and Labor Turnover survey, designed to provide reliable, high-frequency national time-series for the quantities of interest in the flow approach to labor markets, is the source for aggregate statistics which are released monthly on a two-month lag at the national level. BLS has recently begun to release exprimental estimates of JOLTS quantities at the state level as well as by industry.

The Occupational Employment Statistics, released annually by BLS, show employment by occupation and industry at a high level of detail for the national level. Newly released experimental series show a slighly less detailed cross-classification by industry and occupation at the state level. Although less directly relevant to the work outlined above, it is worth noting that this source contains wage and salary information as well.

The Quarterly Census on Employment and Wages, compiled by BLS from administrative data submitted to them by state unemployment insurance agencies, provides a lower-frequency but comprehensive count of covered employment by state and industry. This dataset is used as a universe for both calibration and sampling of BLS employer surveys such as CES. Data at lower levels of geographic detail are released, but subject to heavy suppression. Unlike other, similar sources, the QCEW breaks out employment by size class of employer.

In addition to the Employment Situation report mentioned above, and the state estimates two weeks later, BLS/Census release an anonymized microsample of responses from the Current Population Survey a few weeks after the official statistics. Several organizations, including IPUMS, EPI, and the Atlanta Fed, subsequently release "value-added" edits of the microdata. This data can be extremely valuable, not just for initializing or estimating behavioural equations for individual-level models, but also for estimating cross-tabulations of characteristics for aggregate flow-based models.

Within the Department of Labor, but outside the Bureau of Labor Statistics, the Employment and Training Administration releases weekly data on new and continuing unemployment insurance claims at the state and national levels. At a monthly frequency, partial breakdowns of this data by demographic and employment characteristics are available. Quality may be variable, because the data source is an administrative record maintained for practical purposes rather than a record from a survey fielded to answer research questions.

The Longitudinal Employer-Household Dynamics project, at the U.S. Census Bureau, maintains an extremely rich dataset derived from record linkage of Census Bureau, IRS, and BLS data at the individual household and employer level. Publicly available products include the Quarterly Workforce Indicators, which break down labor market stocks and flows in some detail. Suppression becomes an issue at the county level, but isn't much of a problem as the state level (although state participation is still not 100%). They also pioneered the use of bayesian statistical replicates to release synthetic "on-the-map" data showing commuting information for transport (and other) planning in continuous geography.

The first and second papers of this thesis would rely, in the first instance, on data from the Employment Situation report for information on stocks of employment and population by labor force status and state-level geography. Flows between labor market states, as well as stocks of vacancies would be derived from the JOLTS data. Assuming the integration of annual with quarterly and monthly data goes smoothly, which would be a key advantage of the path signature estimation approach, data from the occupational employment statistics program could be incorporated in order to model higher levels of industry-by-occupation detail (with some assumptions made to model gross flows based on net flows observed from the change in levels). Assuming, similarly, that the estimation methods used are amenable to incorporating missing/censored observations, the QCEW could be used to provide complementary data on industry and geography, as well as size-class of employer. Incorporating information at the highest levels of geographic detail, using LEHD data (or possibly accessing the microdata behind it, subject to an agreement with the Census Bureau) could constitute a longer-run "stretch goal" of the research program outlined for this DPhil thesis.

Agent-Based Models

Given that Agent-Based Models are only likely to be involved in the third paper of this thesis, and the author is very likely to be involved in both using and developing these methods in the intervening time, both at the Topos Institute and in the INET Complexity Economics Group, a discussion of these methods is deferred to future work.

Microsimulation

While an extended treatment of the history and methods of microsimulation modeling will likewise be deferred to future work, it bears mentioning in this context that the individual-level modeling excercise outlined for the third paper will seek to draw on and contribute to the literature and practice of policy microsimulation, which is conceptually closely related to Agent-Based Modeling, but has proceeded largely in parallel to that literature, as outlined in [Cheng, 2020, Ricchiardi, Bronco, and van der Ven, 2023]. The emphasis of this tradition on providing quantitatively accurate analysis of tax and transfer policies dovetails well with the vision outlined in [Sterman, 2018] and the practical, policy-focused goals of the work proposed for this thesis.

Flow Estimation Techniques

The fact that Systems Dynamics models can be represented as ordinary differential equations (see, for example [Baez et. al., 2023] and accompanying software) means that equations determining the flows in the system can be estimated using any techniques developed for estimating the coefficients of ODEs from appropriate data. A "genetic programming" method known as Symbolic Regression has been used for this purpose (see, for example, [Kronberger, Kammerer, and Kommenda, 2020]), as have a number of other computationally-intensive model discovery techniques collectively referred to as "Scientific Machine Learning" by a community of applied mathematicians and programmers who collaborate on an ecosystem of packages in the Julia programming language for handling and estimating such models. A large and growing number of methods for combining differential equations with advanced machine learning methods such as neural networks have been developed in this context (for a list, see https://sciml.ai/citing/). These methods will be explored as part of the second paper, along with methods such as Sparse Identification of Non-linear Dynamical Systems ("SINDy") [Brunton et. al., 2016] and path signature methods [Chevyrev and Kormalitzin, 2016].

In the initial (and simplest) flow model, using national data, there are a little more than 400 monthly observations for each pairwise flow. This should be far more than necessary to estimate a simple linear specification for the flows, but it remains to be seen whether it will be enough to support model selection of nonlinear specifications as outlined in the previous paragraph. This will likely depend on the particular estimation method used, and motivates the exploration in the second paper of the relative merits (as judged by forecasting effectiveness) of the different possible techniques. Each dimension added to the model (such as geography, industry, and occupation) should allow the incorporation of more disaggregated data, and therefore should not dramatically change the type of estimation which is possible.

Applied Category Theory

The connection to following section is somewhat speculative, and the three papers outlined for this thesis do not rely on the correspondence outlined here for their validity or feasibility, but it bears highlighting that several of the methods outlined above, and in particular their implementation in software, have recently been (or are currently being) explored as applications of a unified mathematical framework known as "Applied Category Theory". While it is certainly too early to provide any guarantees to this effect, there is reason to be somewhat confident that these converging projects and strands of the academic literature will facilitate the combined use (and particularly the combined software implementation) of these techniques.

Category Theory is a branch of "pure" mathematics which arose in the mid-20th century as a novel set of techniques for exploring Algebraic Topology. The original impulse behind its development was to formalize the process, ubiquitous in mathematics, of studying the transformations of a mathematical structure which are “natural” in the sense of preserving the essential properties of the structure. The key definition turned out to be two levels “below” the natural transformations themselves, which were conceptualized as structure-preserving mappings between “functors”, which were in turn structure-preserving mappings between “categories”. Categories are defined to consist of a set of objects, an distinguished arrow (or “morphism”) from each object to itself which is taken to be the “identity” map, and a (possibly empty) set of morphisms between any two objects such that the morphisms to and from a particular object can be composed in a logical way to “cut out the middle man”. The identities are (definitionally) neutral with respect to this composition. [Marquis, 2024]

Surprisingly, this three-level structure of categories, functors, and natural transformations turned out to be useful in seemingly unrelated branches of mathematics such as logic and set theory. It has also become extremely popular (if occasionally controversial) in theoretical computer science. [Marquis, 2024] Recently, a group of mathematicians have begun to explore, under the banner of “Applied Category Theory”, a number of applications of category theory to scientific modeling, with a journal, conference series, and some foundational texts [such as Fong and Spivak, 2018] appearing in the last few years.

The use of ideas from Applied Category Theory in software design has already shown some progress in the production of flexible software libraries for producing systems dynamics models [Meadows, Li and Osgood, 2023], and a great deal of effort is now being put into a similar effort aimed at producing software for agent-based modeling [Brown, et. al., 2023]. A particularly promising aspect of the software being developed by these authors, at the Topos Institute and in Dr. Nathaniel Osgood’s Lab at the University of Saskatchewan, is the commitment to designing software that allows a compositional specification of the model. In theory, model builders should be able to compose models in several ways. One would be by identifying some of the stocks and flows of one model with corresponding stocks and flows in another model, but the more obviously useful style of composition (for the present purposes) is what Meadows, Li, and Osgood refer to as “stratification”. The stratification of two models essentially multiplies them together, resulting in a stock for each combination of stocks in the original model. This should allow the modeler to specify different dimensions of a model separately, and combine them using code. By doing so, the modeler should be able to construct, visualize, and estimate much larger and more complex models in a tractable manner. The data structure necessary to perform this sort of stratification, both in Systems Dynamics and in Agent-Based Models, is being developed in the StateCharts.jl package by XiaoYan Li at (https://github.com/AlgebraicJulia/StateCharts.jl). All of these projects are discussed in greater detail at (https://www.algebraicjulia.org/).

Primary Model Inputs and Outputs

The flow-based model estimated from BLS macro time-series for the first and second papers should be able to produce stochastic forecasts of labor market stocks and flows (in particular, unemployment rates) conditional on input data reflecting labor market demand. The stochastic aspect of the forecasts should reflect the estimation error in the flow equations, as well as in the measurement of stocks.

Input data specifying labor market demand could reflect expert-selected scenarios (sourced from the World Bank, or IEA, for example), or else could be derived from a Computable General Equilibrium or Input-Output model forecasting GDP by sector and geography, which could then be translated into a flow of sector and geography-specific vacancies using assumptions (possibly dynamic) on labor intensity by sector. Models such as those used in [Klimek et. al., 2019] and [Pichler et. al., 2022] could provide useful inputs.

Broadly speaking, this modeling effort will take inspiration from the Labour Market model currently under development at INET, which uses an Input-Output framework to translate time-paths of spending estimates in particular sectors into demand estimates for those and upstream sectors [Bucker et. al., 2023]. As mentioned in that paper, there are a large number of possible sources for transition path estimates, and the input scenarios will have to be selected based on a thorough evaluation of the methods an assumptions used to generate them.

Model outputs will be stochastic forecasts of labour market variables including employment, labour market participation, and unemployment (and changes in each of those stocks), along with estimates of the variability of those forecasts which take into account model estimation uncertainty (and may include uncertainty of the input series, when available). If all goes to plan, these estimates will be produced at the level of U.S. States, by industry, occupation, and key demographic characteristics.

Preliminary Chapter Outline

- Introduction - An overview of the three papers with an emphasis on the continuity of methods and subject matter.
- Chapter 1: Empirical Systems Dynamics of Labor Market Flows. An introduction to labour flow modeling and system dynamics methods, as well as results from empirical estimation of a detailed labour flow model on high-frequency data covering the United States.

- Chapter 2: Path Signatures and Competing Methods for Empirical Systems Dynamics. An exploration and analysis of the estimation methods used or considered for use in the preceding chapter.

- Chapter 3: An Agent-Based Perspective on Inequality and Household Finance. This chapter will quantify the connection between income inequality and asset price inflation using an ABM of the dynamics of labour income and household balance sheets.
- Conclusion - A discussion of policy implications of the first and third papers, and methodological recommendations based on the second.
- Appendices - Software products developed for the three papers will be presented in appendices.

Progress to Date

For the first paper, I have largely identified the data and methods which I plan to use, and conducted a literature review (see above). I have also identified the programming language, and in many cases the specific libraries (or at least "ecosystem") that I intend to use. I will be implementing both the modeling and analysis in Julia, a new(ish) programming language designed for the applied mathematics community which has attained a high level of both usability and performance, and is increasingly being adopted by economists and data scientists. In particular, I have identified the StockFlow.jl package, which should enable me to compose smaller stock-flow models into a larger model which can then be expressed (in machine-representable form) as a system of differential equations. This can then be solved, represented, or have its coefficients estimated using the set of methods and Julia libraries associated with the "Scientific Machine Learning" (SciML) project. These libraries already include code for symbolic regression and a large number of methods involving neural networks and more traditional machine learning techniques. I have not yet been able to identify Julia code for path signature/iterated integral methods, so I am looking into either using one of a number of existing libraries written in Python, using cross-language interoperation methods that already exist, or else coding up the desired methods "from scratch" in Julia.

For the individual-level (and household-level) modeling proposed in the third paper, I have identified a number of existing models and modeling frameworks in Python, and at least two frameworks in Julia that may be available on the relevant time horizon [Datseris, Vahdati, and DuBois, 2022, and Brown, et. al., 2023].

I have begun to code the stock-flow labour market model using StockFlow.jl, starting with a "minimum viable" model in which each flow is governed by a single parameter (i.e. constant) flow rate. Model code and output are illustrated below. I have also acquired data from the Bureau of Labor Statistics LABSTAT database for the estimation stage, and implemented a strategy for storing and accessing large amounts of data using SQLite databases. While I have included the JOLTS data, and am preparing to use it for subsequent iterations of the model, it turned out that BLS has released complete CPS-based flow estimates at the national level, which will be sufficient for the proof-of-concept stage of the model.

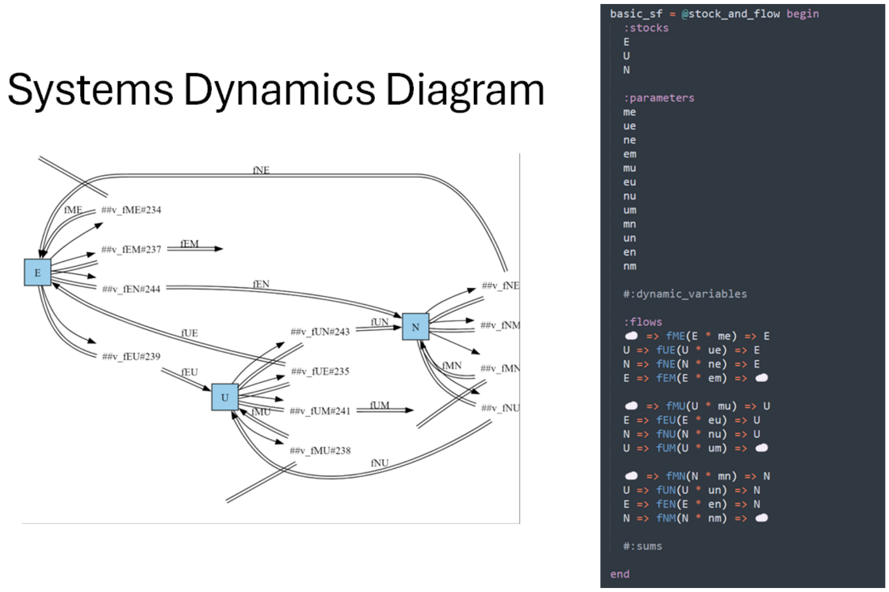

Figure 3. Screenshot of the initial model specification and a program-generated representation of the model's stock-flow structure.

I have also acquired expert judgement data on recession timing at the national level ("official" definitions from the National Bureau of Economic Research) and at the state level (from the Philadelphia Fed), which should help to separate out the estimates of the recovery and recession dynamics.

I've coded a custom loss function for the SciML library which should be able to select optimal parameters for the proof-of-concept model, based on CPS-based flow data from 1990 onward. I'm still debugging the code for this, as the optimisation procedure doesn't (presently) converge. In general, I'm confident that I'll be able to make good progress on the implementation this summer, as I'll be working closely with the authors of a large number of the code libraries which I'm using.

Timetable to Completion

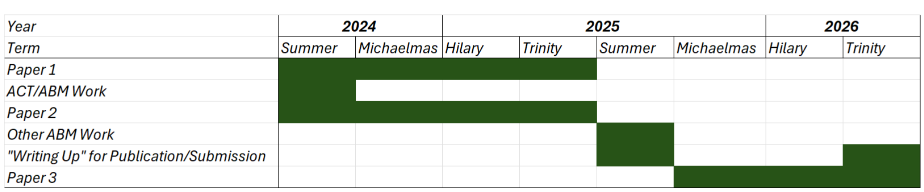

Figure 1: image

My goal is to work concurrently on the first two papers (methods and application), and complete them by the end of my second year, and to complete the third paper in my third and final year of the DPhil. I anticipate devoting some time, particularly during the summers, to getting high-level experience with relevant methods on projects that are related to but perhaps not contained within the papers outlined above. This summer, I will be working with the Topos Institute in Berkeley, CA, on a project in which they are hoping to use ideas from Applied Category Theory to design software for Agent-Based Modeling in Julia (to complement their existing efforts in stock-flow modeling).

At the same time, I have begun to build the model for my first two papers, and anticipate making some progress over the summer. I would expect to complete this model (and the papers based on its methods and results) over the course of the next academic year.

While primarily working on the first two during the next academic year, I would also begin familiarizing myself with one of the ABMs (Macro or Labour) currently under development in the Complexity Group at INET, or else building out a basic labour model based on the previous stock-flow paper and techniques from the Topos Institute project. This would allow me to spend the third year adding housing/household finance capabilities to the model and using it to complete the third paper by the end of my third academic year.

Bibliography

Abraham, Katharine G., John C. Haltiwanger, and Lea E. Rendell. “How Tight Is the US Labor Market?” Brookings Papers on Economic Activity 2020, no. 1 (2020): 97–165.

Ahn, Hie Joo, and James D. Hamilton. “Heterogeneity and Unemployment Dynamics.” Journal of Business & Economic Statistics 38, no. 3 (July 2, 2020): 554–69. https://doi.org/10.1080/07350015.2018.1530116.

Ahn, Hie Joo, Bart Hobijn, and Ayşegül Şahin. “The Dual Us Labor Market Uncovered.” National Bureau of Economic Research, 2023. https://www.nber.org/papers/w31241.

M. Aragón, F et al. (2022). Job displacement costs of phasing out coal. London: Institute for Fiscal Studies. Available at: https://ifs.org.uk/publications/job-displacement-costs-phasing-out-coal (accessed: 7 April 2024).

Autor, David, David Dorn, and Gordon H. Hanson. “On the Persistence of the China Shock.” Working Paper. Working Paper Series. National Bureau of Economic Research, October 2021. https://doi.org/10.3386/w29401.

Baez, John, Xiaoyan Li, Sophie Libkind, Nathaniel D. Osgood, and Evan Patterson. “Compositional Modeling with Stock and Flow Diagrams.” Electronic Proceedings in Theoretical Computer Science 380 (August 7, 2023): 77–96.

https://doi.org/10.4204/EPTCS.380.5.

Barbrook-Johnson, Pete, and Alexandra S. Penn. “System Dynamics.” In Systems Mapping: How to Build and Use Causal Models of Systems, edited by Pete Barbrook-Johnson and Alexandra S. Penn, 113–28. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-031-01919-7_8.

Barlevy, Gadi, R. Jason Faberman, Bart Hobijn, and Ayşegül Şahin. “The Shifting Reasons for Beveridge Curve Shifts.” Journal of Economic Perspectives 38, no. 2 (May 2024): 83–106. https://doi.org/10.1257/jep.38.2.83.

Board of Governors of the Federal Reserve System. 2018. “Speech by Chairman Powell on Monetary Policy in a Changing Economy.” Accessed April 4, 2024. https://www.federalreserve.gov/newsevents/speech/powell20180824a.htm.

Brown, Kristopher, Evan Patterson, Tyler Hanks, and James Fairbanks. “Computational Category-Theoretic Rewriting.” arXiv, March 31, 2023. https://doi.org/10.48550/arXiv.2111.03784.

Brunton, Steven L., Joshua L. Proctor, and J. Nathan Kutz. “Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems.” Proceedings of the National Academy of Sciences 113, no. 15 (April 12, 2016): 3932–37. https://doi.org/10.1073/pnas.1517384113.

Bucker, Joris, R. Maria Del Rio-Chanona, Anton Pichler, Matthew C. Ives, and J. Doyne Farmer. “Employment Dynamics in a Rapid Decarbonization of the Power Sector.” INET Working Paper, November 2023.

Cheng, Chung-Tang. “Guy H. Orcutt’s Engineering Microsimulation to Reengineer Society.” History of Political Economy 52, no. S1 (December 1, 2020): 191–217. https://doi.org/10.1215/00182702-8718000.

Chevyrev, Ilya, and Andrey Kormilitzin. “A Primer on the Signature Method in Machine Learning.” arXiv, March 11, 2016. https://doi.org/10.48550/arXiv.1603.03788.

Cohen, Samuel N., Silvia Lui, Will Malpass, Giulia Mantoan, Lars Nesheim, Áureo de Paula, Andrew Reeves, Craig Scott, Emma Small, and Lingyi Yang. “Nowcasting with Signature Methods.” arXiv, May 17, 2023. https://doi.org/10.48550/arXiv.2305.10256.

Curtis, E. Mark, and Ioana Marinescu. “Green Energy Jobs in the United States: What Are They, and Where Are They?” Environmental and Energy Policy and the Economy 4 (January 1, 2023): 202–37. https://doi.org/10.1086/722677.

Datseris, George, Ali R. Vahdati, and Timothy C. DuBois. “Agents.Jl: A Performant and Feature-Full Agent-Based Modeling Software of Minimal Code Complexity.” SIMULATION, January 5, 2022, 003754972110688. https://doi.org/10.1177/00375497211068820.

del Rio-Chanona, R. Maria, Penny Mealy, Mariano Beguerisse-Díaz, François Lafond, and J. Doyne Farmer. “Occupational Mobility and Automation: A

Data-Driven Network Model.” Journal of The Royal Society Interface 18, no. 174 (January 20, 2021): 20200898. https://doi.org/10.1098/rsif.2020.0898.

Ebrahimy, Ehsan, and Robert Shimer. “Stock–Flow Matching.” Journal of Economic Theory, Search Theory and Applications, 145, no. 4 (July 1, 2010): 1325–53. https://doi.org/10.1016/j.jet.2010.02.012.

Elsby, Michael, Bart Hobijn, Fatih Karahan, Gizem Kos¸ar, and Ays¸egül S¸ahin. “Flow Origins of Labor Force Participation Fluctuations.” AEA Papers and Proceedings 109 (May 1, 2019): 461–64. https://doi.org/10.1257/pandp.20191054.

Fermanian, Adeline. “Functional Linear Regression with Truncated Signatures.” arXiv, June 16, 2022. https://doi.org/10.48550/arXiv.2006.08442.

Farmer, J. Doyne, Cameron Hepburn, Penny Mealy, and Alexander Teytelboym. “A Third Wave in the Economics of Climate Change.” Environmental and Resource Economics 62, no. 2 (October 1, 2015): 329–57. https://doi.org/10.1007/s10640-015-9965-2.

Fong, Brendan, and David I. Spivak. “Seven Sketches in Compositionality: An Invitation to Applied Category Theory.” arXiv, October 12, 2018. https://doi.org/10.48550/arXiv.1803.05316.

Forrester, Jay W. “System Dynamics—a Personal View of the First Fifty Years.” System Dynamics Review 23, no. 2–3 (2007): 345–58. https://doi.org/10.1002/sdr.382.

Hall, Robert E., and Marianna Kudlyak. “Churn and Stability: The Remarkable Heterogeneity of Flows among Employment, Job Search, and Non-Market Activities in the US Population.” Working paper, 2022. https://www.ecb.europa.eu/pub/conferences/shared/pdf/20230921_8th_ARC/2023_ARC_Kudlyak_paper.pd

Hall, Robert E., and Marianna Kudlyak. “Why Has the US Economy Recovered So Consistently from Every Recession in the Past 70 Years?” NBER Macroeconomics Annual 36 (May 2022): 1–55. https://doi.org/10.1086/718588.

Hall, Robert E., and Marianna Kudlyak. “The Active Role of the Natural Rate of Unemployment during Cyclical Recoveries.” Working Paper. Working Paper Series. National Bureau of Economic Research, November 2023. https://doi.org/10.3386/w31848.

Herrington, Gaya. “Update to Limits to Growth: Comparing the World3 Model with Empirical Data.” Journal of Industrial Ecology 25, no. 3 (2021): 614–26. https://doi.org/10.1111/jiec.13084.

Hobijn, Bart, and Ayşegül Şahin. “Maximum Employment and the Participation Cycle.” National Bureau of Economic Research, 2021. https://www.nber.org/papers/w29222.

Jo, Chulsu, Doo Hwan Kim, and Jae Woo Lee. “Forecasting Unemployment and Employment: A System Dynamics Approach.” Technological Forecasting and Social Change 194 (September 1, 2023): 122715. https://doi.org/10.1016/j.techfore.2023.122715.

Klimek, Peter, Sebastian Poledna, and Stefan Thurner. “Quantifying Economic Resilience from Input–Output Susceptibility to Improve Predictions of Economic Growth and Recovery.” Nature Communications 10, no. 1 (April 11, 2019): 1677. https://doi.org/10.1038/s41467-019-09357-w.

Kronberger, Gabriel, Lukas Kammerer, and Michael Kommenda. Identification of Dynamical Systems Using Symbolic Regression. Vol. 12013, 2020. https://doi.org/10.1007/978-3-030-45093-9.

Lyons, Terry, and Andrew D. McLeod. “Signature Methods in Machine Learning.” arXiv, January 26, 2024. https://doi.org/10.48550/arXiv.2206.14674.

J J McCall. Economics of Information and Job Search. The Quarterly Journal of Economics, 84(1):113–126, 1970.

Marquis, Jean-Pierre. “Category Theory.” In The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta and Uri Nodelman, Fall 2023. Metaphysics Research Lab, Stanford University, 2023. https://plato.stanford.edu/archives/fall2023/entries/category-theory/.

Meadows, Donella H., Dennis L. Meadows, Jørgen Randers, and William W. Behrens Iii. “The Limits to Growth - Club of Rome,” January 1, 1972. https://policycommons.net/artifacts/1529440/the-limits-to-growth/2219251/.

Meadows, Nicholas, Xiaoyan Li, and Nathaniel D. Osgood. “Hierarchical and Upstream-Downstream Composition of Stock and Flow Models.” arXiv, May 3, 2023. http://arxiv.org/abs/2305.02136.

Mian, Atif R., Ludwig Straub, and Amir Sufi. “The Saving Glut of the Rich.” Working Paper. Working Paper Series. National Bureau of Economic Research, April 2020. https://doi.org/10.3386/w26941.

Morrill, James, Adeline Fermanian, Patrick Kidger, and Terry Lyons. “A Generalised Signature Method for Multivariate Time Series Feature Extraction.” arXiv, February 6, 2021. https://doi.org/10.48550/arXiv.2006.00873.

Mortensen, Dale T. “Markets with Search Friction and the DMP Model.” The American Economic Review 101, no. 4 (2011): 1073–91.

Nishimura, Hirokazu. “Synthetic Differential Geometry of Chen’s Iterated Integrals.” arXiv, November 9, 2014. https://doi.org/10.48550/arXiv.1406.0075.

Pichler, A., Pangallo, M., del Rio-Chanona, R.M., Lafond, F. & Farmer, J.D. (2022). ‘Forecasting the propagation of pandemic shocks with a dynamic input-output model’. Journal of Economic Dynamics and Control, Volume 144, November 2022, 104527

Radosavljevic, Sonja, Thomas Banitz, Volker Grimm, Lars-Göran Johansson, Emilie Lindkvist, Maja Schlüter, and Petri Ylikoski. “Dynamical Systems

Modeling for Structural Understanding of Social-Ecological Systems: A Primer." Ecological Complexity 56 (2023): 101052.

Radzicki, Michael J. "System Dynamics and Its Contribution to Economics and Economic Modeling." In Encyclopedia of Complexity and Systems Science, edited by Robert A. Meyers, 8990–9000. New York, NY: Springer, 2009. https://doi.org/10.1007/978-0-387-30440-3_539.

Richiardi, Matteo, Patryk Bronka, and Justin van de Ven. Back to the future: Agent-based modelling and dynamic microsimulation. No. CEMPA8/23. Centre for Microsimulation and Policy Analysis at the Institute for Social and Economic Research, 2023.

Solow, Robert M. "Is the End of the World at Hand?." Challenge 16, no. 1 (1973): 39-50.

Sterman, John. "System Dynamics at Sixty: The Path Forward." System Dynamics Review 34, no. 1–2 (2018): 5–47. https://doi.org/10.1002/sdr.1601.

Stern, Nicholas, Joseph Stiglitz, and Charlotte Taylor. "The Economics of Immense Risk, Urgent Action and Radical Change: Towards New Approaches to the Economics of Climate Change." Journal of Economic Methodology 29, no. 3 (July 3, 2022): 181–216. https://doi.org/10.1080/1350178X.2022.2040740.

Stiglitz, Joseph E. "Where Modern Macroeconomics Went Wrong." Oxford Review of Economic Policy 34, no. 1–2 (January 5, 2018): 70–106. https://doi.org/10.1093/oxrep/grx057.