A Big Data Approach to Optimal Taxation

Richard W. Evans, Kenneth L. Judd, and Kerk L. Phillips

July 8, 2014

Introduction

Big data refers to any repository of data that is either large enough or complex enough that distributed and parallel input and output approaches must be used (see White, 2012, p. 3). Einav and Levin (2013) discuss the new opportunities in economics using big data, although they focus primarily on searching for important patterns in existing datasets. Our line of research is a different application of the big data approach in that we use it as a solution method to theoretical models rather than an empirical method. Our solution method is ideally suited for dealing with models with nonconvex optimization problems, a large degree of heterogeneity, rich policy structure, potential model uncertainty, and potential policy objective uncertainty.

Our big data approach involves generating a large database of optimal agent behavior for an efficiently chosen sample of points in both the space of agent types and the space of tax policies. We then use big data techniques to access this database in order to calculate social welfare and total government revenue for each point in the tax policy space. Next, we eliminate all tax policies that are strictly dominated by any other policy in terms of both social welfare and total tax revenue. This leaves us with a tax policy frontier in terms of social welfare and total tax revenue. The last step is to iteratively refine our search for optimal tax policies around the points remaining on the frontier and repeat the deletion step of strictly dominated policies. The result is a close approximation of the continuous function of tax policies that traces out the frontier of optimal social welfare and tax revenue possibilities.

Tax Model Taxonomy

The universe of economic tax policy models is broad, varied, and quite disparate. There is no grand unifying theory of macroeconomics (or microeconomics) that explains even a majority of the phenomena observed throughout the economy. The main reason for this diversity and disparity is the complexity of the system in which the economic phenomena are generated. The current state of macroeconomics is to use separate models for particular sets of research or policy questions, which models abstract from the characteristics of the economy less relevant to the question and focus on the most important characteristics influencing the economic question at hand.

A skeleton and a crash test dummy are two separate models or approximations of the human body. They look very different from each other because they are used to answer different questions about the body. In the same way, the different economic models used to answer tax policy questions can often look starkly different from each other.

We identify seven key dimensions on which these tax models differ and within which each of the models can be categorized. We the two extremes of each of these dimensions in order from least mathematically and computationally complex to most complex. Each of these dimensions represent tradeoffs that that the modeler or researcher must make in the face of binding analytical and computational constraints.

1. Partial equilibrium vs. general equilibrium
2. Static vs. dynamic
3. Deterministic vs. stochastic
4. Reduced form vs. structural
5. Linear vs. nonlinear
6. Positive vs. normative
7. Low heterogeneity vs. high heterogeneity

No one model is at the high end of each of the above categories. For example, the Dynamic Stochastic General Equilibrium (DSGE) models of the New Dynamic Public Finance (NDPF) research are at the high end of categories 1 through 6, but necessarily must limit the degree of heterogeneity in those models (low on dimension 7).[1] New Keynesian DSGE models, such as Smets and Wouters (2007), are also high on dimensions 1 through 6, but are more often solved with linear methods (medium dimension 5), have more explicit economic structures although more of those are ad hoc frictions. Because of their structural complexity, Keynesian DSGE models can handle medium levels of heterogeneity (dimension 7) at most.

Input-Output models are large systems of equations (often in the hundreds) that represent firm and industry behavior. These models are ? equilibrium, dynamic, deterministic, reduced form linear, positive forecasting machines. They incorporate large scale firm heterogeneity, but are light on household modeling.

The Congressional Budget Office (CBO) uses the CBO Long Term Model (CBOLT) to forecast the Social Security Trust Fund balance 75 years into the future. This model is a large-scale overlapping generations microsimulation model that is general equilibrium, dynamic, structural, and has a large degree of heterogeneity. However, this model must be deterministic in order to be tractable. Further, parameter weights are adjusted to make the CBOLT macroeconomic outcomes match up with the CBO's 10-year baseline macroeconomic forecast values, which come from a separate and

[1] See Kocherlakota (2010) for an overview of the New Dynamic Public Finance.

different CBO macroeconomic model. The CBOLT model is mostly used for positive economics—although it has the individual utility function structures necessary for normative analysis—because solving for optimal policy in that environment is currently not analytically or computationally tractable.

JCT microsimulation models used to score (forecast revenue changes) proposed legislation are partial equilibrium (static scoring), deterministic, extremely reduced form (down to elasticities), and are only used for positive analysis. However, these models are dynamic and incorporate a large amount of heterogeneity.

Our Big Data Approach

Our "big data" approach to studying optimal tax policy focuses on structural models with nonlinear solutions and a high degree of heterogeneity that can be used for normative welfare analysis (high on dimensions 4 through 7). As such, our current studies using this method use models that are partial equilibrium, static, and deterministic (low no dimensions 1 through 3).

In particular, we focus on optimal policy problems with a high degree of heterogeneity that generate a nonconvex problem for the policy maker to solve. A number of seminal papers in the commodity tax literature made strong assumptions on the demand functions of individuals in order to attain aggregation results, thereby keeping the policy maker's optimization problem convex.[2] However, Deaton (1977, p.310) concedes:

> The result rests on strong simplifying assumptions in order to avoid the complexity of the general case; in particular, consumer bahaviour has been restricted by the use of linear Engel curves and by permitting only very limited substitution between commodities. Relaxing either of these could alter quite fundamentally the nature of the empirical results.

The following is a simplified, but general, example of our big data approach. Suppose that the economy is populated by a large number of households that differ from one another along multiple dimensions. Let those dimensions be indexed by θ_i, where θ_1 is one type of household and θ_2 is another type of household. Assume that these individuals must make some economic choices $\boldsymbol{x}$—which could denote things like consumption of different goods, how much to work, how much to save, etc.—taking as given some tax schedule that they face $\boldsymbol{\tau}$. If we knew the tax schedule $\boldsymbol{\tau}$ they faced, and the households' problem were well specified, then we could solve for the households' optimal decisions $\boldsymbol{x}(\theta_i, \boldsymbol{\tau})$, utility at those optimal decisions $u(\theta_i, \boldsymbol{\tau})$, and taxes paid at those optimal decisions $r(\theta_i, \boldsymbol{\tau})$—all as functions of the individuals type θ_i and the tax schedule $\boldsymbol{\tau}$.

Our big data approach is to choose a grid over individual types $\boldsymbol{\theta}$, essentially choosing the number of different types of households. We then choose a grid of possible tax policies to consider $\boldsymbol{\tau}$. For each tax policy $\boldsymbol{\tau}_j$ in the grid of all possible

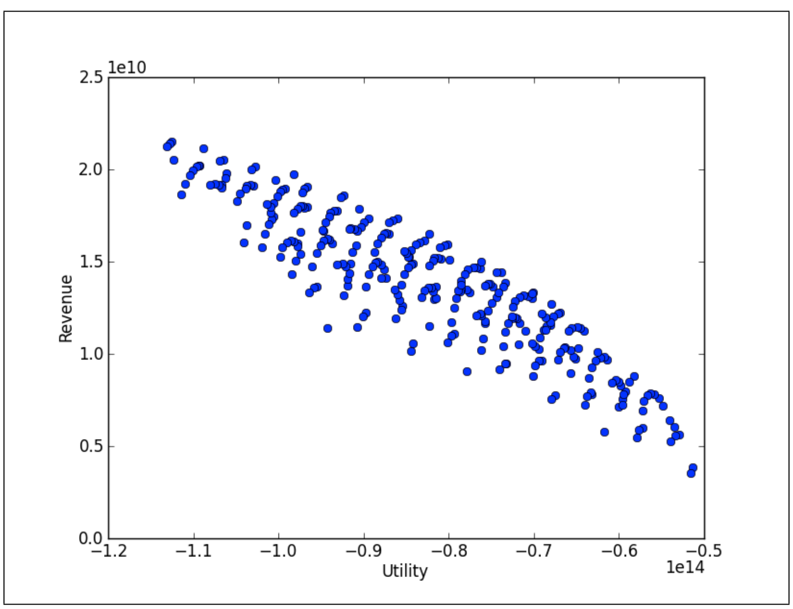
[2] See Diamond and Mirrlees (1971a,b) and Deaton (1977).

tax policies $\boldsymbol{\tau}$, we can compute all the individual optimal responses $\boldsymbol{x}(\theta_i, \boldsymbol{\tau}_j)$, each of their utilities $u(\theta_i, \boldsymbol{\tau}_j)$, and the tax revenues they pay $r(\theta_i, \boldsymbol{\tau}_j)$. We store these responses in a large database.

The size of the database is what requires big data techniques. For example, suppose there were two dimensions of heterogeneity among household types $\boldsymbol{\theta}$—ability and dislike for labor, for example—and there were 10 different ability types and 10 different disutilities of labor. That would mean that we had 100 different types of individuals. Now consider the tax policy space. Suppose there were two goods, each of which having a sales tax rate τ_1 and τ_2. If we wanted to look at 10 different tax rates for each tax $\tau_1, \tau_2 = 0\%, 10\%, 20\%...90\%$, that would be 100 different combinations of tax schedules. For each of the 100 possible tax schedules we would compute the individual responses, utilities and tax revenues. That equals $10,000 = 100 \times 100$ individual problems to solve. For realistic versions of this problem, the database often grows to sizes that must be stored across multiple hard drives, thereby requiring the use of big data techniques.

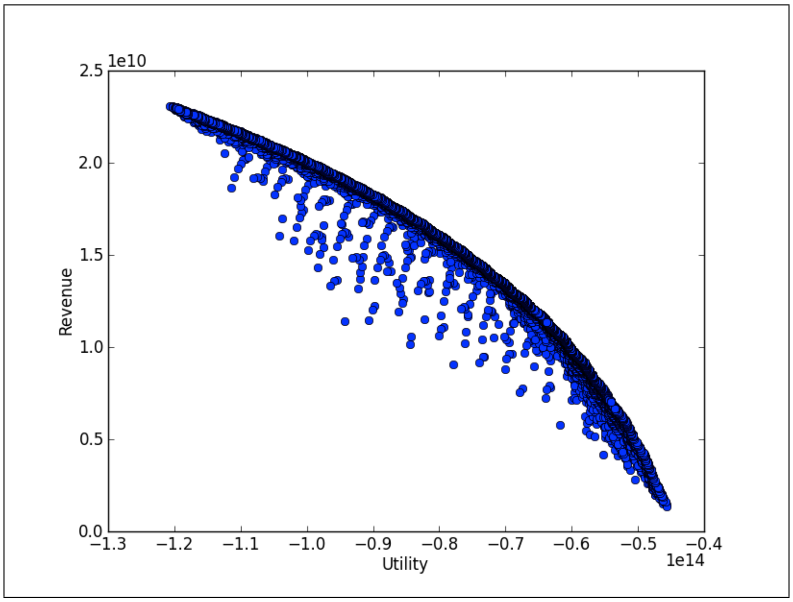
Once we have generated this database grid over individual types and tax policies, the hard part is done. We then decide the distribution over household types, which is making an assumption about the population distribution over types. Once we know the population distribution, we can add up all the individual utilities for a given tax policy $\boldsymbol{\tau}_j$ to get a total utility value U_j for each possible tax policy. We can likewise add up all the individual tax payments for a given tax policy to get total tax revenues for a given tax policy R_j. Figure 1 shows different (U_j, R_j) combinations for all the possible tax policies in tax space $\boldsymbol{\tau}$.

Figure 1: Total utility-total revenue (U_j, R_j) combinations for all tax schedules $\boldsymbol{\tau}_j$



From Figure 1, it is clear that the tax policies associated with points that are strictly dominated in terms of (U_j, R_j) are not optimal.[3] We then highlight the tax policies that are not strictly dominated that form the frontier of total utility and total revenue as shown in Figure 2. We then choose points around the points on the frontier and repeat the process of deleting strictly dominated strategies and highlighting the frontier. We call this refinement. Figure 2 shows a refined frontier.

Figure 2: Total utility-total revenue (U_j, R_j) frontier and refinement



The resulting total utility-total revenue frontier in Figure 2 corresponds to all the tax schedules that maximize total utility for a given level of total revenue. In other words, for a desired level of total revenue on the y-axis, the the point on the frontier in a straight leftward horizontal line tells us the tax policy that maximizes social welfare. The optimization problem that determines each point on the frontier is a nonconvex problem and, therefore, requires these specialized solution methods. Our big data solution method solves for the entire frontier at once. This is valuable, especially if a policymaker has some uncertainty about what total revenue targe they want to hit. We solve for the entire frontier at once and can, therefore, accommodate uncertainty about the total revenue constraint.

This exposition of the big data approach to tax policy can be applied to a large number of problems. In the paragraphs below, we detail some applications that we are currently working on as well as some intended future directions.

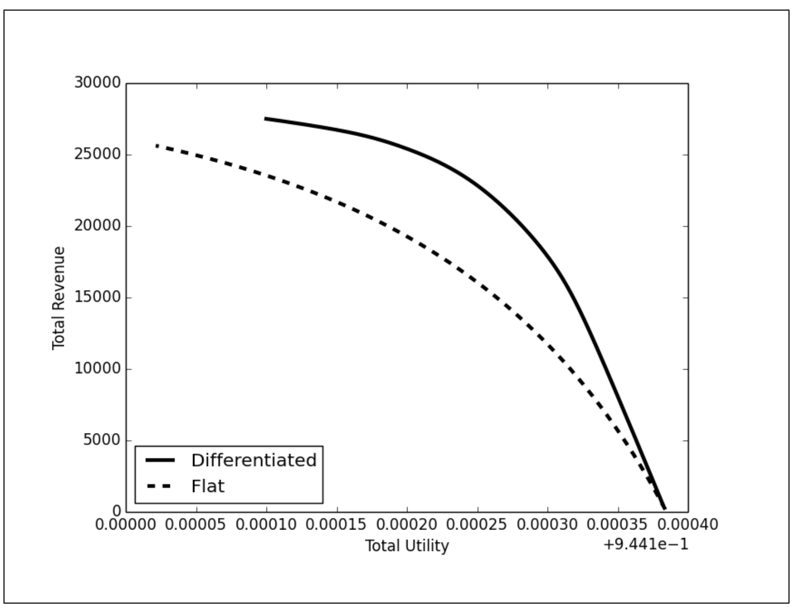
[3] Strictly dominated points in Figure 1 are points for which there exists another point that has both higher total utility U_j and higher total tax revenue R_j.

Sales taxation example

In our most recent paper, Baker et al. (2014) use this framework to study some key questions about optimal sales tax policy. Calibrating heterogeneity among the households in the model to the U.S. income distribution, estimated U.S. price markups, and consumption category shares by consumer income, we study the difference between an optimally differentiated sales tax across goods and a flat tax. We also quantify the loss from excluding the broad services category from sales taxation, as is the case in the U.S. and many other countries.

Figure 3 shows two total utility-total revenue frontiers, each point of which corresponds to the tax policy that provides a combination of societal utility and total tax revenue that is not strictly dominated by any other sales tax schedule. The solid line represents the sales tax schedules on the total utility-total revenue frontier for optimally differentiated sales tax policy. The dashed line represents the total utility-total revenue combinations resulting from optimally chosen flat commodity tax rates. Figure 4 shows the optimal tax rates on the eight different goods in our model corresponding to points along each of the frontiers in Figure 3.

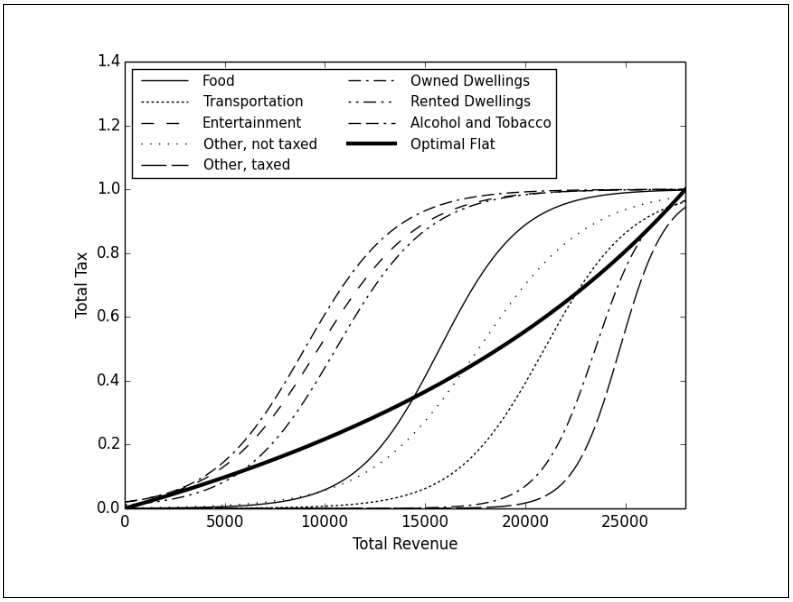
Figure 3: Total utility-revenue frontiers for optimal differentiated tax versus optimal flat tax



As expected, solid line total utility-total revenue frontier for the optimally differentiated sales tax schedules is everywhere above the frontier for the optimal flat tax. This is, roughly speaking, because the optimally differentiated tax places higher sales tax rates on goods with higher demand elasticities,[4] while the optimal flat tax

[4]This confirms an early finding by Ramsey (1927).

Figure 4: Optimal tax rates for good i for different levels of total revenue



is required to tax all goods at the same rate.

However, we find that the revenue loss of an optimal flat tax regime relative to the optimally differentiated tax regime might not be that large. The percent loss in total revenue for a given total utility level for each of the frontiers in Figure 3 hits a maximum of just over 30 percent. This loss might not be that large if one considers the information and enforcement requirements and costs associated with an optimally differentiated sales tax system. We interpret this as evidence that a broad-based flat tax system might be a reasonable option for fundamental tax reform.

We also studied the effect of exempting the broad services category of goods from sales taxation, as is currently the case in the U.S. as well as many other countries. Figures 5 and 6 show that the total revenue loss in exempting services from sales taxation is approximately constant at about 2 percent. This small loss is due to increased sales taxes on low elasticity goods to compensate for the revenue loss from exempted services. The result holds for both the optimally differentiated sales tax schedule and the optimal flat tax.

Figure 5: Total utility-revenue frontiers for optimal differentiated tax versus optimal differentiated tax with services exempted

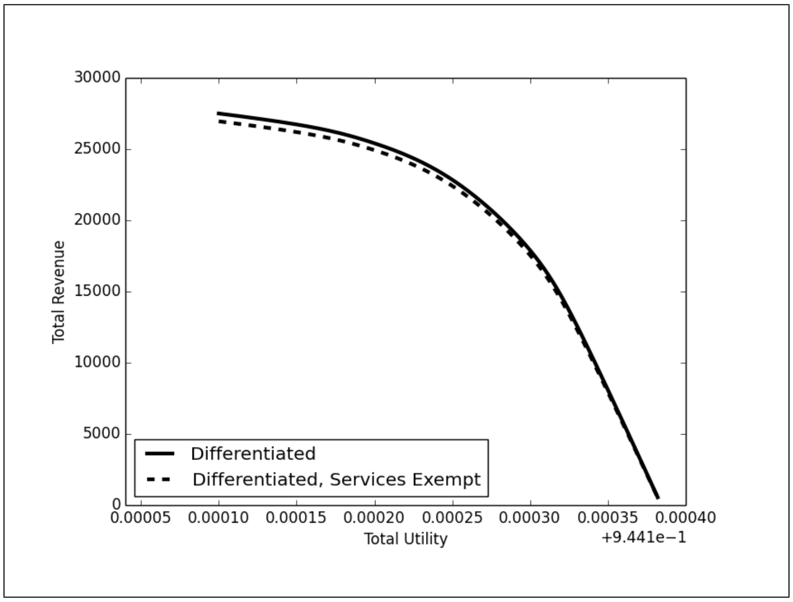
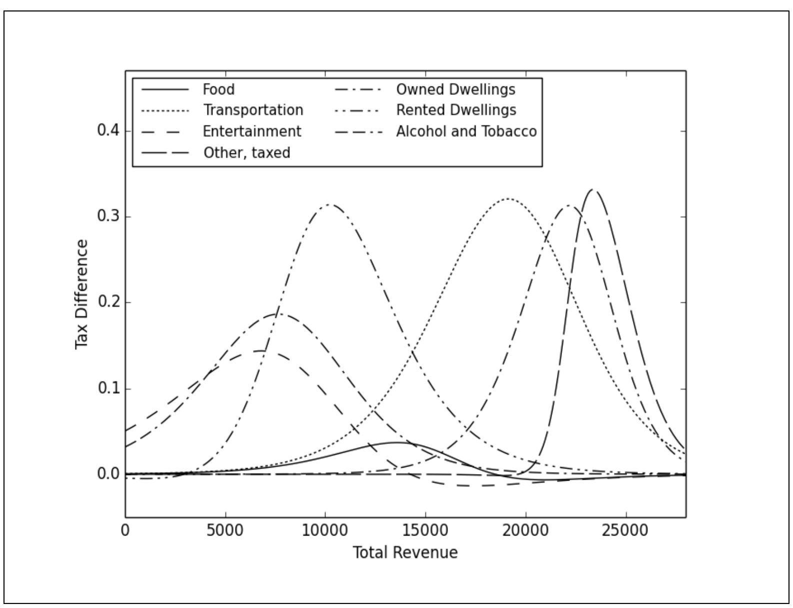


Figure 6: Difference in other seven optimally differentiated tax rates including non-taxed services minus tax rates exempting services

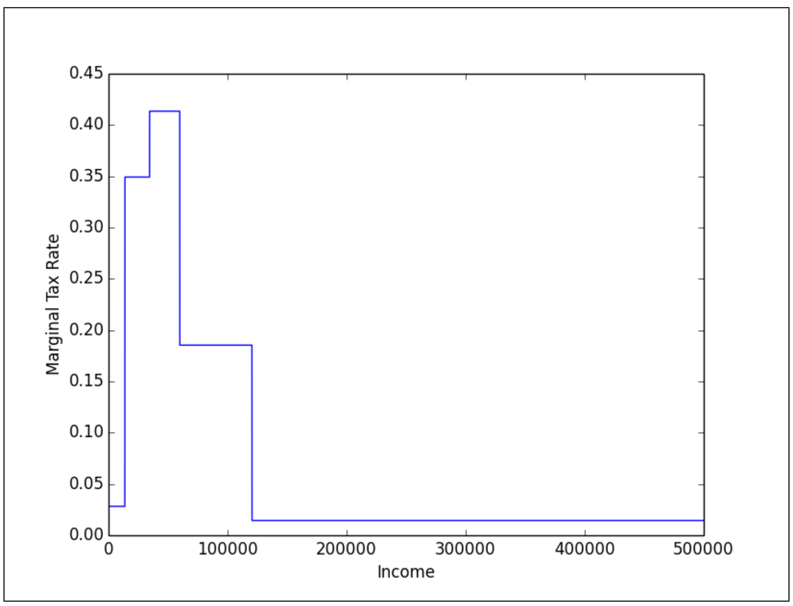


Income taxation example

As another application of our big data optimal taxation research line, we have recently begun to study optimal income tax schedules. Judd and Su (2006) provide evidence that when more dimensions of heterogeneity among households are modeled, results such as negative marginal tax rates on high income households can be observed. However, they were not able to compute robust examples in their framework. With our big data approach, we can study optimal piecewise linear marginal income tax schedules in the face of multidimensional household heterogeneity.

In our preliminary tests as shown in Figure 7, we find that the optimal piecewise linear marginal tax rate schedule with only five tax rates matches the results predicted by Mirrlees (1971) in their functional analysis version of the problem. Our marginal tax rate begin close to zero and end close to zero. Our next step is to ramp up the dimensions of heterogeneity and see how the results change.

Figure 7: Optimal piecewise linear marginal tax rate schedule with one-dimensional heterogeneity and five tax rates



Conclusion

Beyond the sales tax and income tax examples, other extensions of this modeling approach include making the individual's problem dynamic and studying the optimal tax rates in general equilibrium. Both extensions require a significant increase in computational complexity. However, current supercomputing resources and computational methods make this big data approach ideal for answering optimal tax questions.

References

Baker, Christian, Jeremy Bejarano, Richard W. Evans, Kenneth L. Judd, and Kerk L. Phillips, "A Big Data Approach to Optimal Sales Taxation," NBER Working Paper No. 20130, National Bureau of Economic Research May 2014.

Deaton, Angus, "Equity, Efficiency, and the Structure of Indirect Taxation," *Journal of Public Economics*, December 1977, *8* (3), 299–312.

Diamond, Peter A. and James A. Mirrlees, "Optimal Taxation and Public Production I: Production Efficiency," *American Economic Review*, March 1971, *61* (1), 8–27.

_ and _ , "Optimal Taxation and Public Production II: Tax Rules," *American Economic Review*, June 1971, *61* (3, Part 1 of 2), 261–278.

Einav, Liran and Jonathan D. Levin, "The Data Revolution and Economic Analysis," NBER Working Paper No. 19035, National Bureau of Economic Research May 2013.

Judd, Kenneth L. and Che-Lin Su, "Optimal Income Taxation with Multidimensional Taxpayer Types," November 2006. mimeo.

Kocherlakota, Narayana R., *The New Dynamic Public Finance*, Princeton University Press, 2010.

Mirrlees, J. A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, April 1971, *38* (114), 175–208.

Ramsey, Frank P., "A Contribution to the Theory of Taxation," *Economic Journal*, March 1927, *37* (145), 47–61.

Smets, Frank and Rafael Wouters, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, June 2007, *97* (3), 586–606.

White, Tom, *Hadoop: The Definitive Guide*, O'Reilly Media, 2012.