Summary of Essays on Bayesian Econometrics for Big Data

by

Kenichiro McAlinn

Keio University, Graduate School of Economics

Chapter 1
Introduction

With the increased availability of large-scale data in many fields, including economics, finance, and marketing, the need for econometric tools to cope with the challenges associated with dealing with big data has increased in recent years. While the term "Big Data" has been used with different definitions, such as data that cannot be stored on memory being the original definition, here we define it as large-scale data that cannot be effectively estimated using traditional methods, e.g., OLS, Bayesian regression with non-informative priors etc. For example, one prominent case of this is the $p >> n$ problem, where the number of parameters exceeds the number of data. In this case, ordinary econometric tools fail in properly estimating the model, leading to improper forecasts, inference, and decision making. Although many methods have been developed to deal with these challenges, especially in the field of statistical/machine learning, not all developments are done with economic datasets in mind. To this end, this dissertation develops methodological and computational tools or modeling, inference, forecasting, and decision making for big data.

There are several benefits in analyzing big data for economic analysis. One prevailing practice in economics and finance is to model at the aggregate level. For example, while there are hundreds or thousands of assets that can be analyzed, many analysis in the literature analyze the aggregate indices. Although modeling the aggregate has its benefits, such as having lesser noise than its individual components, the results might be misleading in terms of its generalization. As such, there are many examples where a property found in the aggregate level does not appear to exist (or at least not as strong as the aggregate level) in its components. Analyzing a large set of data, and developing models for its individual components, greatly increases the economic insight gained from the analysis.

Another benefit is its practical utility. With a large dataset at hand, the economist should use all of the data in order to learn about the complex structure of the evolving economy, or information in order to make better decisions. Utilizing big data, thus, allows the economist to make a fully informed analysis without constraint or bias (e.g. selection bias). The effectiveness of big data tools have been shown in many fields, but with limited success in economics. Developing econometric tools to address the key points in the field to fully utilize large-scale data is therefore critical.

Chapter 2

Non-Linear Leverage Effects: Evidence From A Large Cross-Section of Stock Returns

The estimation, inference, and prediction of volatility is one of the most crucial aspects in analyzing data with variability. In the field of finance and economics, volatility of financial assets has been investigated with great scrutiny to further the understanding of the mechanics and structure of price movement. One aspect of volatility that has gathered special interest is the correlation between an asset's return and its volatility; coined the *leverage effect*. In particular, for decision making involving predictions, this correlation is critical, as knowing how today's change will effect tomorrow's risk is simply necessary for most sequential decision problems, especially under considerable shocks. It is often claimed that this correlation is negative, implying that a negative (positive) shock to an asset's return results in an increase (decrease) in its volatility. Thus, changing decisions accordingly based on predictions of increased or decreased volatility, implied by the previous shock.

This phenomenon is intuitive, as we can expect– and often observe– that an asset under distress exhibits more variability and uncertainty compared to an asset that is stable or increasing in price. The term *leverage* refers to an economic interpretation given by Black (1976) and Christie (1982). They state that, when an asset's price declines, the company's relative debt increases, making the balance sheet *leveraged*, resulting in the company being riskier and therefore its market value more volatile (see Bekaert and Wu 2000, for example, for different interpretations and comparisons of the leverage effect). Though only a hypothesis, this explanation has held weight in the field and the effect is widely believed to exist, with supporting evidence from examining major stock indices (Nelson 1991; Glosten et al. 1993; Dumas et al. 1998, for ARCH-type models and Jacquier et al. 1994; Harvey and Shephard 1996; West and Harrison 1997; Jacquier et al. 2004; Yu 2005; Omori et al. 2007; Nakajima and Omori 2009; Asai and McAleer 2011; Asai et al. 2012; Nakajima and Omori 2012; Takahashi et al. 2013; Shirota et al. 2014, for SV-type models). However, contrary to consensus, the lack of empirical evidence of the effect from individual stocks is paradoxical; with most stocks exhibiting zero or very weak correlation between asset returns and volatility. This is troublesome for decision makers wishing to exploit this structure for decision making, since

mischaracterization of this correlation can lead to considerable loss in utility.

We postulate that this is caused by the simple, but almost universal, representation of the correlation: Most volatility models in the literature, basic or advanced, assume that the relationship between an asset's return and its volatility is linear, even though many advances have been made on other aspects of the model. However, it is counter-intuitive to think that a large shock in return effects volatility with the same linear relationship as small daily fluctuations. This notion has promoted research in considering more complex leverage effects. For example, Hansen et al. (2012) introduced a more general form of the leverage effect by using a leverage function within the GARCH framework. In the context of stochastic volatility (SV) models, there has been no advances in this direction, even though SV models are known to outperform ARCH-type models due to its flexibility in capturing traits seen in asset returns (Geweke 1994; Fridman and Harris 1998; Kim et al. 1998). The advances are hindered, partly, due to the computational complexity SV models entail, as it requires complex Markov chain Monte Carlo (MCMC) methods that are hard to sample and tune.

We respond to this movement by extending the SV model to include a leverage function in the form of a Hermite polynomial to examine the nonlinear dynamics of the correlation between an asset's return and its volatility. To achieve this, we develop an effective Bayesian computation method using sequential Monte Carlo (SMC) by developing and extending the particle learning method of Carvalho et al. (2010), enabling estimation of the parameters of interest in a fast, efficient, and online manner that would have been previously near impossible. With the new model and algorithm, we are able to examine and analyze the leverage effect over a large number of equity assets and over time, and find strong evidence for the leverage effect where it is unobserved, or weak, under the simple linear representation, with improved predictive performance.

Chapter 3

Parallel Computing for Large-Scale On-Line Estimation

The state space model (SSM) has been one of the indispensable tools for time series analysis and optimal control for decades. Although the archetypal SSM is linear and Gaussian, the literature of more general non-linear and non-Gaussian SSMs has been rapidly growing in the last two decades. For lack of an analytically tractable way to estimate the general SSM, numerous approximation methods have been proposed. Among them, arguably the most widely applied method is particle filtering (Gordon et al. 1993; Kitagawa 1996). Particle filtering is a type of sequential Monte Carlo method in which the integrals we need to evaluate for filtering are approximated by the Monte Carlo integration. To improve numerical accuracy and stability of the particle filtering algorithm, various extensions, such as the auxiliary particle filter (Pitt and Shephard 1999a), have been proposed, and still actively studied by many researchers. For SSMs with unknown parameters, Kitagawa (1998) proposed a self-organizing state space modeling approach in which the unknown parameters are regarded as a subset of the state variables and the joint posterior distribution of the parameters and the state variables is evaluated with a particle filtering algorithm. Other particle filtering methods that can simultaneously estimate parameters have been proposed by Liu and West (2001), Storvik (2002), Fearnhead (2002), Polson et al. (2008), Johannes and Polson (2008), Johannes et al. (2008), Carvalho et al. (2010), just to name a few. These particle filtering methods that estimate state variables and parameters simultaneously are often called particle learning methods in the literature. Although the effectiveness of particle filtering methods have been proven through many different applications (see Montemerlo et al. (2003), Zou and Chakrabarty (2007), Mihaylova et al. (2008), Chai and Yang (2007), Lopes and Tsay (2011), and Dukic et al. (2012) among others), it is offset by the fact that it is a time-consuming technique. Some practitioners still shy away from using it in their applications because of this despite its benefit.

This attitude toward particle filtering would be changed by the latest technology: Parallel computing. As we will discuss in Section 2, some parts of the particle filtering procedure are ready to be executed simultaneously on many processors in a parallel computing environment. In light

of inexpensive parallel processing devices such as GPGPUs[1] (general purpose graphics processing units) available to the general public, more and more researchers start to jump on the bandwagon of parallel computing. Suchard et al. (2010) and Lee et al. (2010) reviewed general attempts at parallelization of Bayesian estimation methods. Durham and Geweke (2011) implemented a sequential Monte Carlo method on the GPU and applied it to complex nonlinear dynamic models, which are numerically intractable even for the Markov chain Monte Carlo method.

As for parallelization of particle filtering, a few researches (see Montemayor et al. (2004), Bolic et al. (2005), Maskell et al. (2006), Hendeby et al. (2007), Hendeby et al. (2010), Chao et al. (2010), Gong et al. (2012), for example) have been reported, though the field is still in a very early stage. However, all of these state of the art methods are either 1) simple implementations onto parallel devices, 2) modifies the algorithm in a way that introduces additional estimation errors, or 3) depends on device-specific functionalities that would make it inapplicable for other devices. For example, Lee et al. (2010) and Durham and Geweke (2011) are both parallel implementations of the particle filter, however, the resampling step for both implementations are computed sequentially and thus they are not fully parallel algorithms. Hendeby et al. (2010) and Gong et al. (2012), on the other hand, use device-specific functionalities of the GPU to parallelize the resampling step that cannot be implemented in other devices. One method that has been used is what is called local resampling (Chao et al. 2010), which breaks up the resampling step in to several blocks and sequentially resample within that block. This method is obviously not a fully parallel algorithm as, while the computationally burden is lessened, it requires sequential computation within blocks, and thus not exploiting the full power of the parallel framework. A fully parallel algorithm for particle filters have yet to be developed, to the best of the authors' knowledge.

In developing parallel algorithms, with particularly GPUs in mind, there are a few bottlenecks one should avoid. First, processing sequential algorithms on the GPU can be inefficient because of the GPU's device memory architecture and its lack of clock speed compared to the CPU. Roughly speaking, a GPU has two types of memory: memory assigned to each core and memory shared

[1] A high-performance GPU (graphics processing unit) was originally developed for displaying high-resolution 2D/3D graphics necessary in video games and computer-aided design. Because a GPU is designed with a massive number of processor cores to conduct single-instruction multiple-data (SIMD) processing, it has been regarded as an attractive platform of parallel computing and researchers started to use it for high-performance computing. As GPU manufacturers try to take advantage of this opportunity, it has evolved into a more computation-oriented device called GPGPU. Nowadays almost all GPUs have more or less capabilities for parallel computing, so the distinction between GPUs and GPGPUs are blurred.

by all cores. Access to the core-linked memory is fast while access to the shared memory takes more time. Generally, one should try as much to keep all calculations on each core without any large-scale communications among cores. The second bottleneck is that it is time-consuming to transfer memory between the host memory, which the CPU uses, and the device memory, which the GPU uses. In other words, the bandwidth between the GPU's device memory and the CPU's host memory is very narrow. We can see that a fully parallel algorithm defined above would be ideal for GPU devices as it would, automatically and without manipulation, be able to calculate everything within the GPU and without bottlenecks.

With these bottlenecks in mind, we have developed a new parallel resampling algorithm to complete the first fully parallel algorithm that computes the full cycle of the particle filtering algorithm in a massively and fully parallel manner. This includes the computing of the likelihood for each particle, constructing the cumulative distribution function (CDF) for resampling, resampling the particles with the CDF, and propagating new particles for the next cycle. By keeping all of our computations within the GPU and avoiding all memory transfer between the GPU and the CPU during the execution of the particle filtering algorithm, we exploit the great benefits of parallel computing to the fullest while avoiding its short comings, especially on the GPU. While we designed our algorithm with GPUs in mind, since our parallel algorithm is a fully parallel algorithm, it can be easily implemented on other parallel computing devices.

In order to compare our new parallel algorithm with conventional sequential algorithms, we conduct a Monte Carlo experiment in which we apply the competing particle learning algorithms to estimate a simple state space model (stochastic trend with noise model) and record the execution time of each algorithm. The results show that our parallel algorithm on the GPU is far superior to the conventional sequential algorithm on the CPU by around $30\times$. Focusing only on the resampling step, we have achieved a speed up of around $10\times$, which considering the sequential nature of the algorithm, is a significant improvement.

Chapter 4
Dynamic Sparse Factor Analysis

The premise of dynamic factor analysis (DFA) is fairly straightforward: there are unobservable commonalities in the variation of observable time series, which can be exploited for interpretation, forecasting, and decision making. Dating back to, at least, Burns and Mitchell (1947), the fundamental idea that a small number of indices drive co-movements of many time series has found plentiful empirical support across a wide range of applications including economics (Stock and Watson 2002a; Bai and Ng 2002; Bernanke et al. 2005; Baumeister et al. ????; Cheng et al. 2016), finance (Diebold and Nerlove 1989; Aguilar et al. 1998; Pitt and Shephard 1999b; Aguilar and West 2000; Carvalho et al. 2011), and ecology (Zuur et al. 2003), to name just a few. More notably, in their seminal work on DFA, Sargent et al. (1977) showed that two dynamic factors could explain a large fraction of the variance of U.S. quarterly macroeconomic variables. Motivated by a similar (but significantly larger) application, we develop scalable Bayesian DFA methodology and deploy it to glean insights into the hidden drivers of the U.S. macroeconomy before, during and after the Great Recession.

With large-scale cross sectional data becoming readily available, the need for developing scalable and reliable tools adept at capturing complex latent dynamics have spurred in both statistics and econometrics (Beyeler and Kaufmann 2016; Kaufmann and Schumacher 2017; Fruehwirth-Schnatter and Lopes 2018; Nakajima et al. 2017). While "dynamic factor models have been the main big data tool used over the past 15 years by empirical macroeconomists" (Stock and Watson 2016), there are remaining methodological challenges. It is now commonly agreed that high-dimensional inference can hardly be formalized and executed without any sparsity assumptions. The fundamental goal of our research is to facilitate sparsity *discovery* (i.e. data-informed sparsity), when in fact present. In doing so, we keep in mind three main pillars that we regard as essential for building a stable foundation for sparse factor modeling.

Firstly, the latent factor loadings should account for time-varying patterns of sparsity. In (macro-)economics and finance, the sequentially observed variables may go through multiple periods of shocks, expansions, and contractions (Hamilton 1989). It is thus expected that the underlying latent structure changes over time– either gradually or suddenly– where some factors might be

active at all times, while others only at certain times. For example, in our empirical analysis we find that certain factors exert influence on some series only during a crisis and later permeate through different components of the economy as the shock spreads. Dynamic sparsity plays a very compelling role in capturing and characterizing such dynamics. Recent developments in sparse factor analysis reflect this direction of interest (West 2003; Carvalho et al. 2008; Yoshida and West 2010; Lopes et al. 2010). More recently, Nakajima and West (2013c) deployed the latent threshold approach of Nakajima and West (2013a) in order to induce zero loadings dynamically over time. Our methodological contribution builds on this development, but poses far less practical limitations on the dimensionality of the data and far less constraints on identification.

Related to the previous point is the question of selecting the number of factors. This modeling choice is traditionally determined by a combination of *a priori* knowledge, a visual inspection of the scree plot (Onatski 2009), and/or information criteria (Bai and Ng 2002; Hallin and Liska 2007). In the presence of model uncertainty, the Bayesian approach affords the opportunity to assign a probabilistic blanket over various models. Bayesian non-parametric approaches have been considered for estimating the factor dimensionality using sparsity inducing priors (Bhattacharya and Dunson 2011; Rockova and George 2016). The added difficulty stemming from time series data, however, is that the number of factors *may change over time*. Despite plentiful empirical evidence for this behavior in macroeconomic data (Bai and Ng 2002), the majority of existing DFA tools treat the number of factors as fixed over time. As a remedy, we turn to dynamic sparsity as a compass for determining the number of factors without necessarily committing to one fixed number ahead of time.

The third essential requirement is accounting for structural instabilities over time with time-varying loadings and/or factors. One seemingly simple solution has been to deploy rolling/extending window approaches to obtain pseudo-dynamic loadings. These estimates, however, lack any supporting probabilistic structure that would induce smoothness and/or capture sudden dynamics. Recent DFA developments (Del Negro and Otrok 2008; Nakajima and West 2013a) have treated both the factors and loadings as stochastic and dynamic. Adopting this point of view, we blend smoothness with sparsity via Dynamic Spike-and-Slab (DSS) priors on factor loadings (Rockova and McAlinn 2017). This prior regards factor loadings as arising from a mixture of two states: an inactive state represented by very small loadings and an active state represented by smoothly

evolving large loadings. The mixing weights between these two states themselves are time-varying, reflecting past information to prevent from erratic regime switching. The DSS priors allow latent factors to effectively, and smoothly, appear or disappear from each series, tracking the evolution of sparsity over time.

In this work, we develop methodology for sparse dynamic factor analysis that is built on the three foundational principles mentioned above. Using this methodology, we examine a large-scale balanced panel of macroeconomic indices that span multiple corners of the U.S. economy from 2001 to 2015. Our method helps understand how the economy evolves over time and how shocks affect its individual components. In particular, examining the latent factor structure before, during, and after the Great Recession, we obtain insights into the channels of dependencies and we assess permanence of structural changes.

To ensure that our implementation scales with large datasets, we propose an EM algorithm for MAP estimation that recovers evolving sparse latent structures in a fast and potent manner. As the EM algorithm finds a likely sparse structure, it does not require strong identification constraints that would be needed for MCMC simulation. While interpretation can be achieved with ex-post rotations (Bai and Ng 2013; Kaufmann and Schumacher 2017), here we deploy rotations to sparsity *inside* the EM algorithm along the lines of Rockova and George (2016) to (a) accelerate convergence and (b) obtain better oriented sparse solutions.

Chapter 5
Large-Scale Predictive Regression

The increasing availability of large datasets, both in terms of the number of variables and the number of observations, combined with the recent advancements in the field of econometrics, statistics, and machine learning, have spurred the interest in predictive models with many explanatory variables, both in finance and economics.[2] As not all predictors are necessarily relevant, decision makers often pre-select the most important candidate explanatory variables by appealing to economic theories, existing empirical literature, and their own heuristic arguments. Nevertheless, a decision maker is often still left with tens– if not hundreds– of sensible predictors that may possibly provide useful information about the future behavior of quantities of interest. However, the out-of-sample performance of standard techniques, such as ordinary least squares, maximum likelihood, or Bayesian inference with uninformative priors tends to deteriorate as the dimensionality of the data increases, which is the well known curse of dimensionality.[3]

Confronted with a large set of predictors, two main classes of models became popular, even standard, within the regression framework. *Sparse* modeling focus on the selection of a sub-set of variables with the highest predictive power out of a large set of predictors, and discard those with the least relevance. LASSO-type regularizations are by far the most used in both research and practice. Regularized models take a large number of predictors and introduce penalization to discipline the model space. Similarly, in the Bayesian literature, a prominent example is the spike-and-slab prior proposed by George and McCulloch (1993), which introduced variable selection through a data-augmentation approach. A second class of models fall under the heading of *dense* modeling; this is based on the assumption that, a priori, all variables could bring useful information for prediction, although the impact of some of these might be small. As a result, the statistical features of a large set of predictors are assumed to be captured by a much smaller set of common latent components, which could be either static or dynamic. Factor analysis is a clear example of

[2] See, e.g., Elliott and Timmermann (2004), Timmermann (2004), Bai and Ng (2010), Rapach et al. (2010), Billio et al. (2013), Manzan (2015), Pettenuzzo and Ravazzolo (2016), Harvey et al. (2016), Giannone et al. (2017), and McAlinn and West (2017), just to cite a few.

[3] Even with a moderate number of predictors the empirical investigation of all possible model combinations could rapidly become infeasible. For instance, for a moderate size linear regression with $p = 30$ regressors, investigating the whole set of possible features combinations would require estimating $2^{30} \approx 1.07$ billion regression models.

dense statistical modeling, which is highly popular in applied macroeconomics (see, e.g., Stock and Watson (2002b) and De Mol et al. (2008) and the references therein).

Both of these approaches entail either an implicit or explicit reduction of the model space. The intention is to arbitrarily lower model complexity to balance bias and variance, in order to potentially minimize predictive losses. For instance, in LASSO-type shrinkage estimators, increasing the tuning parameter (i.e. increasing shrinkage) leads to a higher bias, thus using cross-validation aims to balance the bias-variance tradeoff by adjusting the tuning parameter. Similarly, in factor models, the optimal number of latent common components is chosen by using information criteria to reduce the variance by reducing the model dimensionality at the cost of increasing the bias (see, e.g., Bai and Ng (2002)). In addition, for economic and financial decision making, in particular, these dimension reduction techniques always lead to a decrease in consistent interpretability, something that might be critical for policy makers, analysts, and investors.

In this paper, we propose a novel class of data-rich predictive synthesis techniques and contribute to the literature on predictive modeling and decision making with large datasets. We take a significantly different approach towards the bias-variance tradeoff by breaking a large dimensional problem into a set of small dimensional ones. More specifically, we retain all of the information available and *decouple* a large predictive regression model into a set of smaller regressions constructed by clustering the set of regressors into J different groups, each one containing fewer regressors than the whole, according to their economic meaning or some quantitative clustering. Rather than assuming a priori the existence of a sparse structure or few latent common components, we retain all of the information by estimating J different predictive densities– separately and sequentially– one for each group of predictors, and *recouple* them dynamically to generate aggregate predictive densities for the quantity of interest. By decoupling a large predictive regression model into smaller, less complex regressions, we keep the aggregate model variance low while sequentially learning and correcting for the misspecification bias that characterize each group. As this is the case, the recoupling step benefits from biased models, as long as the bias has a signal that can be learned. This flips the bias-variance tradeoff around, exploiting the weakness of low complexity models to an advantage in the recoupling step, therefore improving the out-of-sample predictive performance.

We implement the proposed the methodology, which we call decouple-recouple synthesis (DRS),

and explore both its econometric underpinnings and economic gains on both a macroeconomic and a finance application. More specifically, in the first application we test the performance of our decouple-recouple approach to forecast the one- and three-, and twelve-month ahead annual inflation rate in the U.S. over the period 1986/1 to 2015/12, a context of topical interest (see, e.g., Cogley and Sargent (2005), Primiceri (2005), Stock and Watson (2007), Koop et al. (2010), and Nakajima and West (2013b), among others). The set of monthly macroeconomic predictors consists of an updated version of the Stock and Watson macroeconomic panel available at the Federal Reserve Bank of St.Louis. Details on the construction of the dataset can be found in McCracken and Ng (2016). The second application relates to forecasting monthly year-on-year total excess returns across different industries in the U.S. from 1970/1 to 2015/12, based on a large set of both industry-specific and aggregate predictors. The predictors have been chosen from previous academic studies and existing economic theory (see, e.g., Goyal and Welch (2008) and Rapach et al. (2010)).

Forecasting accuracy is assessed in a statistical sense based on two different out-of-sample performance metrics. We report as a main performance metric the Log Predictive Density Ratio (LPDR), at forecast horizon k and across time indices t. In addition, although our main focus is on density forecasts, we also report the Root Mean Squared Forecast Error (RMSFE), which captures the forecast optimality for a mean squared utility. Irrespective of the performance evaluation metric, our decouple-recouple model synthesis scheme emerges as the best for forecasting the yearly total excess returns across different industries. The differences in the LPDRs are stark and clearly shows a performance gap in favor of DRS.

As far as the out-of-sample economic performance is concerned, we run a battery of tests based on a power-utility representative investor with moderate risk aversion. The comparison is conducted for the unconstrained as well as short-sales constrained investor at monthly horizons, for the entire sample. We find that our DRS strategy results in a higher CER (relative to an investor that uses the historical mean as forecast) of more than 150 basis points per year, on average across sectors. Consistent with the predictive accuracy results, we generally find that the DRS strategy produces higher CER improvements than the competing specifications, both with and without short-sales portfolio constraints. In addition, we show that DRS allows to reach a higher CER also on a "per-period" basis, which suggests that there are economically important gains for a power

utility investor.

References

Aguilar, O., Huerta, G., Prado, R., and West, M. (1998), "Bayesian inference on latent structure in time series," *Bayesian Statistics*, 6, 1–16.

Aguilar, O. and West, M. (2000), "Bayesian dynamic factor models and portfolio allocation," *Journal of Business & Economic Statistics*, 18, 338–357.

Asai, M. and McAleer, M. (2011), "Alternative asymmetric stochastic volatility models," *Econometric Reviews*, 30, 548–564.

Asai, M., McAleer, M., and Medeiros, M. C. (2012), "Modelling and forecasting noisy realized volatility," *Computational Statistics & Data Analysis*, 56, 217–230.

Bai, J. and Ng, S. (2002), "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221.

— (2010), "Instrumental variable estimation in a data rich environment," *Econometric Theory*, 26, 1577–1606.

— (2013), "Principal components estimation and identification of static factors," *Journal of Econometrics*, 176, 18–29.

Baumeister, C., Liu, P., and Mumtaz, H. (????), "Changes in the transmission of monetary policy: Evidence from a time-varying factor-augmented VAR," *Working Paper No. 401, Bank of England*.

Bekaert, G. and Wu, G. (2000), "Asymmetric volatility and risk in equity markets," *Review of Financial Studies*, 13, 1–42.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005), "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach," *The Quarterly journal of Economics*, 120, 387–422.

Beyeler, S. and Kaufmann, S. (2016), "Factor augmented VAR revisited: A sparse dynamic factor model approach," Tech. rep., Working Paper, Study Center Gerzensee.

Bhattacharya, A. and Dunson, D. B. (2011), "Sparse Bayesian infinite factor models," *Biometrika*, 291–306.

Billio, M., Casarin, R., Ravazzolo, F., and van Dijk, H. K. (2013), "Time-varying combinations of predictive densities using nonlinear filtering," *Journal of Econometrics*, 177, 213–232.

Black, F. (1976), "Studies of stock price volatility changes," *In: Proceedings of the 1976 Meetings of the American Statistical Association*, 171–181.

Bolic, M., Djuric, P. M., and Hong, S. (2005), "Resampling algorithm and architectures for distributed particle filters," *IEEE Transactions on Signal Processing*, 53, 2442–2450.

Burns, A. F. and Mitchell, W. C. (1947), *Measuring Business Cycles*, The National Bureau of Economic Research.

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., and West, M. (2008), "High-dimensional sparse factor modeling: applications in gene expression genomics," *Journal of the American Statistical Association*, 103, 1438–1456.

Carvalho, C. M., Johannes, A. M., Lopes, H. F., and Polson, N. G. (2010), "Particle learning and smoothing," *Statistical Science*, 25, 88–106.

Carvalho, C. M., Lopes, H. F., and Aguilar, O. (2011), "Dynamic stock selection strategies: A structured factor model framework," *Bayesian Statistics*, 9, 1–21.

Chai, X. and Yang, Q. (2007), "Reducing the calibration effort for probabilistic indoor location estimation," *IEEE Transactions on Mobile Computing*, 6, 649–662.

Chao, M., Chu, C., Chao, C., and Wu, A. (2010), "Efficient parallelized particle filter design on CUDA," *IEEE Workshop on Signal Processing Systems*, 299–304.

Cheng, X., Liao, Z., and Schorfheide, F. (2016), "Shrinkage estimation of high-dimensional factor models with structural instabilities," *The Review of Economic Studies*, 83, 1511–1543.

Christie, A. A. (1982), "The stochastic behavior of common stock variances: Value, leverage and interest rate effects," *Journal of Financial Economics*, 10, 407–432.

Cogley, T. and Sargent, T. J. (2005), "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S." *Review of Economic Dynamics*, 8, 262–302.

De Mol, C., Giannone, D., and Reichlin, L. (2008), "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" *Journal of Econometrics*, 146, 318–328.

Del Negro, M. and Otrok, C. (2008), "Dynamic factor models with time-varying parameters: measuring changes in international business cycles," *FRB of New York Staff Report No.326*.

Diebold, F. X. and Nerlove, M. (1989), "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model," *Journal of Applied Econometrics*, 4, 1–21.

Dukic, V. M., Lopes, H. F., and Polson, N. G. (2012), "Tracking epidemics with Google flu trends data and a state-space SEIR model," *J. Amer. Statist. Assoc.*, 107.

Dumas, B., Fleming, J., and Whaley, R. (1998), "Implied volatility functions: empirical tests," *Journal of Finance*, 53, 2059–2106.

Durham, G. B. and Geweke, G. (2011), "Massively parallel sequential Monte Carlo for Bayesian inference," *Working paper*.

Elliott, G. and Timmermann, A. (2004), "Optimal forecast combinations under general loss functions and forecast error distributions," *Journal of Econometrics*, 122, 47–79.

Fearnhead, P. (2002), "Markov chain Monte Carlo, sufficient statistics, and particle filters," *J. Comput. Graph. Statist.*, 11, 848–862.

Fridman, M. and Harris, L. (1998), "A maximum likelihood approach for non-Gaussian stochastic volatility models," *Journal of Business and Economics Statistics*, 16, 284–291.

Fruehwirth-Schnatter, S. and Lopes, H. F. (2018), "Sparse Bayesian Factor Analysis when the Number of Factors is Unknown," *arXiv:1804.04231*.

George, E. I. and McCulloch, R. E. (1993), "Variable selection via Gibbs sampling," *Journal of the American Statistical Association*, 88, 881–889.

Geweke, J. (1994), "Bayesian comparison of econometric models," *Working Paper, Federal Reserve Bank of Minneapolis Research Department*.

Giannone, D., Lenza, M., and Primiceri, G. (2017), "Economic predictions with big data: The illusion of sparsity," *Working Paper*.

Glosten, L., Jagannathan, R., and Runkle, D. (1993), "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48, 1779–1801.

Gong, P., Basciftci, Y. O., and Ozguner, F. (2012), "A Parallel Resampling Algorithm for Particle Filtering on Shared-Memory Architectures," *IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum*, 1477–1483.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993), "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEEE Proceedings-F*, 140, 107–113.

Goyal, A. and Welch, I. (2008), "A comprehensive look at the empirical performance of equity premium prediction," *The Review of Financial Studies*, 21, 1455–1508.

Hallin, M. and Liska, R. (2007), "Determining the number of factors in the general dynamic factor model," *Journal of the American Statistical Association*, 102, 603–617.

Hamilton, J. D. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica: Journal of the Econometric Society*, 357–384.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012), "Realized GARCH: a Joint Model For Returns and Realized Mesures of Volatility," *Journal of Applied Econometrics*, 27, 877–906.

Harvey, A. C. and Shephard, N. (1996), "Estimation of an asymmetric stochastic volatility model for asset returns," *Journal of Business & Economic Statistics*, 14, 429–434.

Harvey, C. R., Liu, Y., and Zhu, H. (2016), "... and the cross-section of expected returns," *The Review of Financial Studies*, 29, 5–68.

Hendeby, G., Hol, J. D., Karlsson, R., and Gustafsson, F. (2007), "A Graphics Processing Unit Implementation of the Particle Filter," *Proceedings of the 15th European Statistical Signal Processing Conference*, 1639–1643.

Hendeby, G., Karlsson, R., and Gustafsson, F. (2010), "Particle Filtering: the Need for Speed," *Journal on Advances in Signal Processing*, 22.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371–389.

— (2004), "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 122, 185–212.

Johannes, M. and Polson, N. G. (2008), "Exact Particle Filtering and Learning," *Working paper*, Univ. Chicago Booth School of Business.

Johannes, M., Polson, N. G., and Yae, S. M. (2008), "Non-linear Filtering and Learning," *Working paper*, Univ. Chicago Booth School of Business.

Kaufmann, S. and Schumacher, C. (2017), "Identifying relevant and irrelevant variables in sparse factor models," *Journal of Applied Econometrics*, 32, 1123–1144.

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Review of Economic Studies*, 65, 361–393.

Kitagawa, G. (1996), "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *J. Comput. Graph. Statist.*, 5, 1–25.

— (1998), "A Self-Organizing State-space Model," *J. Amer. Statist. Assoc.*, 93, 1203–1215.

Koop, G., Korobilis, D., et al. (2010), "Bayesian multivariate time series methods for empirical Macroeconomics," *Foundations and Trends® in Econometrics*, 3, 267–358.

Lee, A., Yau, C., Giles, M. B., Doucet, A., and Holmes, C. C. (2010), "On the Utility of Graphics Cards to Perform Massively Parallel Simulation of Advanced Monte Carlo Methods," *J. Comput. Graph. Statist.*, 19, 769–789.

Liu, J. and West, M. (2001), "Combined Parameters and State Estimation in Simulation-based Filtering," *Sequential Monte Carlo Methods in Practice*, (A. Doucet, N. de Freitas and N. Gordon, eds.), 197–223.

Lopes, H. F., McCulloch, R., and Tsay, R. (2010), "Cholesky stochastic volatility," *Unpublished Technical Report, University of Chicago, Booth Business School*, 2.

Lopes, H. F. and Tsay, R. S. (2011), "Particle Filters and Bayesian Inference in Financial Econometrics," *Journal of Forecasting*, 30, 168–209.

Manzan, S. (2015), "Forecasting the distribution of economic variables in a data-rich environment," *Journal of Business & Economic Statistics*, 33, 144–164.

Maskell, S., Alun-Jones, B., and Macleod, M. (2006), "A Single Instruction Multiple Data Particle Filter," *Proceedings of Nonlinear Statistical Signal Processing Workshop*.

McAlinn, K. and West, M. (2017), "Dynamic Bayesian predictive synthesis in time series forecasting," *Journal of Econometrics*, Forthcoming.

McCracken, M. W. and Ng, S. (2016), "FRED-MD: A monthly database for macroeconomic research," *Journal of Business & Economic Statistics*, 34, 574–589.

Mihaylova, L., Angelova, D., Honary, S., Bull, D. R., N., C. C., and Ristic, B. (2008), "Mobility Tracking in Cellular Networks using Particle Filtering," *IEEE Transactions on Wireless Communications*, 6, 3589–3599.

Montemayor, A. S., Pantrigo, J. J., Sanchez, A., and Fernandez, F. (2004), "Particle Filter on GPUs for Real Time Tracking," *Proceedings of the International Conference on Computer Graphics and Interactive Techniques*.

Montemerlo, M., Thrun, S., Koller, D., and Wegbreit, B. (2003), "FastSLAM 2.0: an Improved Particle Filtering Algorithm for Simultaneous Localization and Mapping that Probably Converges," *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 1151–1157.

Nakajima, J. and Omori, Y. (2009), "Leverage, heavy-tails and correlated jumps in stochastic volatility models," *Computational Statistics & Data Analysis*, 53, 2335–2353.

— (2012), "Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Students t-distribution," *Computational Statistics & Data Analysis*, 56, 3690–3704.

Nakajima, J. and West, M. (2013a), "Bayesian analysis of latent threshold dynamic models," *Journal of Business & Economic Statistics*, 31, 151–164.

— (2013b), "Bayesian analysis of latent threshold dynamic models," *Journal of Business & Economic Statistics*, 31, 151–164.

— (2013c), "Dynamic factor volatility modeling: A Bayesian latent threshold approach," *Journal of Financial Econometrics*, 11, 116–153.

Nakajima, J., West, M., et al. (2017), "Dynamics & sparsity in latent threshold factor models: A study in multivariate EEG signal processing," *Brazilian Journal of Probability and Statistics*, 31, 701–731.

Nelson, D. (1991), "Conditional heteroskedasticity in asset pricing: A new approach," *Econometrica*, 59, 347–370.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007), "Stochastic volatility with leverage: fast likelihood inference," *Journal of Econometrics*, 140, 425–449.

Onatski, A. (2009), "Testing hypotheses about the number of factors in large factor models," *Econometrica*, 77, 1447–1479.

Pettenuzzo, D. and Ravazzolo, F. (2016), "Optimal portfolio choice under decision-based model combinations," *Journal of Applied Econometrics*, 31, 1312–1332.

Pitt, M. and Shephard, N. (1999a), "Filtering via Simulation: Auxiliary Particle Filters," *J. Amer. Statist. Assoc.*, 94, 590–599.

— (1999b), "Time varying covariances: a factor stochastic volatility approach," *Bayesian Statistics*, 6, 547–570.

Polson, N. G., Stroud, J. R., and Müller, P. (2008), "Practical Filtering with Sequential Parameter Learning," *J. Roy. Statist. Soc. Ser. B*, 70, 413–428.

Primiceri, G. E. (2005), "Time varying structural vector autoregressions and monetary policy," *Review of Economic Studies*, 72, 821–852.

Rapach, D., Strauss, J., and Zhou, G. (2010), "Out-of-sample equity prediction: Combination forecasts and links to the real economy," *The Review of Financial Studies*, 23, 822–862.

Rockova, V. and George, E. I. (2016), "Fast Bayesian factor analysis via automatic rotations to sparsity," *Journal of the American Statistical Association*, 111, 1608–1622.

Rockova, V. and McAlinn, K. (2017), "Dynamic Variable Selection with Spike-and-Slab Process Priors," *Booth School of Business Technical Report*.

Sargent, T. J., Sims, C. A., et al. (1977), "Business cycle modeling without pretending to have too much a priori economic theory," *New methods in Business cycle research*, 1, 145–168.

Shirota, S., Hizu, T., and Omori, Y. (2014), "Realized stochastic volatility with leverage and long memory," *Computational Statistics & Data Analysis*, 76, 618–641.

Stock, J. H. and Watson, M. W. (2002a), "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

— (2002b), "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

— (2007), "Why has US inflation become harder to forecast?" *Journal of Money, Credit and Banking*, 39, 3–33.

— (2016), "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in Macroeconomics," in *Handbook of Macroeconomics*, Elsevier, vol. 2, pp. 415–525.

Storvik, G. (2002), "Particle Filters in State Space Models with the Presence of Unknown Static Parameters," *IEEE Trans. Signal Process*, 50, 281–289.

Suchard, M. A., Wang, Q., Chan, C., Frelinger, J., Cron, A., and West, M. (2010), "Understanding GPU Programming for Statistical Computation: Studies in Massively Parallel Massive Mixtures," *J. Comput. Graph. Statist.*, 19, 419–438.

Takahashi, M., Omori, Y., and Watanabe, T. (2013), "News impact curve for stochastic volatility models," *Economics Letters*, 120, 130–134.

Timmermann, A. (2004), "Forecast combinations," in *Handbook of Economic Forecasting*, eds. Elliott, G., Granger, C. W. J., and Timmermann, A., North Holland, vol. 1, chap. 4, pp. 135–196.

West, M. (2003), "Bayesian factor regression models in the "large p, small n" paradigm," in *Bayesian Statistics 7*, Oxford University Press, pp. 723–732.

West, M. and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, Springer, 2nd ed.

Yoshida, R. and West, M. (2010), "Bayesian learning in sparse graphical factor models via variational mean-field annealing," *Journal of Machine Learning Research*, 11, 1771–1798.

Yu, J. (2005), "On leverage in a stochastic volatility model," *Journal of Econometrics*, 127, 165–178.

Zou, Y. and Chakrabarty, K. (2007), "Distributed Mobility Management for Target Tracking in Mobile Sensor Networks," *IEEE Transactions on Mobile Computing*, 6, 872–887.

Zuur, A. F., Fryer, R., Jolliffe, I., Dekker, R., and Beukema, J. (2003), "Estimating common trends in multivariate time series using dynamic factor analysis," *Environmetrics*, 14, 665–685.