Title
Essays in Macroeconometrics

Author
LUSOMPA, AMAZE

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Essays in Macroeconometrics

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Amaze Lusompa

Dissertation Committee:
Professor Eric Swanson, Chair
Professor William Branch
Associate Professor Ivan Jeliazkov
Professor Fabio Milani

2020

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1106401, the Federal Reserve Bank of San Francisco's Thomas J. Sargent Dissertation Fellowship, and the Federal Reserve Bank of Boston under the American Economic Association Summer Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation, the Federal Reserve Bank of San Francisco, or the Federal Reserve Bank of Boston.

VITA

Amaze Lusompa

EDUCATION

Doctor of Philosophy in Economics **2020**
University of California, Irvine *Irvine, California*

Master's of Science in Statistical and Economic Modeling **2015**
Duke University *Durham, North Carolina*

Bachelor of Science in Economics **2013**
North Carolina Agricultural and Technical State University *Greensboro, North Carolina*

RESEARCH EXPERIENCE

Graduate Research Assistant **2018–2020**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2018–2020**
University of California, Irvine *Irvine, California*

ABSTRACT OF THE DISSERTATION

Essays in Macroeconometrics

By

Amaze Lusompa

Doctor of Philosophy in Economics

University of California, Irvine, 2020

Professor Eric Swanson, Chair

My dissertation consists of 3 chapters that cover two major themes. The first theme is to provide more accurate, robust, and efficient statistical methods to estimate the impact shocks have on the economy. The second theme is estimating the time-varying and nonlinear impact shocks have on the economy.

What is the impact of a stimulus package? What is the impact of the Federal Reserve increasing the fed funds rate? What are the sources of business cycle fluctuations? These are some of the most important questions in macroeconomics, and to answer them it is often necessary to estimate impulse responses. Over the past decade, Local Projections have been increasingly used to estimate impulse responses. Local Projections are linear regressions that estimate impulse responses directly. The first chapter of my dissertation, "Local Projections, Autocorrelation, and Efficiency", introduces a more accurate and robust method to estimate impulse responses using Local Projection. These methods can be employed using Bayesian or frequentist methods. I also extend Local Projections to be able to estimate time-varying impulse responses.

The second chapter of my dissertation, "Identifying Government Spending Shocks, Fiscal Foresight, and Time-Varying Parameters", focuses on estimating the government spending multiplier, and testing the multiplier for time variation and nonlinearity. Ramey (2011, 2019)

document the wide range of estimates for multipliers in both empirical studies and dynamic simulations. This chapter makes several contributions to the estimation of fiscal multipliers. First, this chapter points out that the current identification schemes for estimating the multiplier are not identifying exogenous changes in government spending, which has caused previous estimates of the multiplier to be biased. I also show that the way the impulse responses are estimated cause additional bias in the multiplier estimates.

The third chapter of my dissertation, "A Kalman Filter Test for Structural Instability", focuses on testing parameter instability in regression models. Researchers are often interested in whether there is parameter instability in regression models. There are several studies that show a significant amount of macroeconomic and financial time series exhibit parameter instability These changes can occur for many reasons such as policy changes, technological evolution, changing economic conditions, etc. There are a multitude of methods available to test for parameter instability, but it has been demonstrated that the most popular tests for parameter instability do not perform adequately. That is, even when there is parameter instability, these tests have trouble detecting the instability. In this chapter, I propose a more powerful test to detect parameter instability.

Chapter 1

Local Projections, Autocorrelation, and Efficiency

Vector Autoregressions (VARs) were proposed in Sims (1980) as an alternative to the large scale simultaneous equation models of the time. Since then, VARs have been a major tool used in empirical macroeconomic analysis, primarily being used for causal analysis and forecasting through the estimation of impulse response functions. In a seminal paper, Jordà (2005) argued that impulse response functions could be estimated directly using linear regressions called Local Projections (LP) and that LP are more robust to model misspecification than VARs.[1,2] LP have been growing in popularity ever since, and the two methods often give different results when applied to the same problem (Ramey, 2016; Nakamura and Steinsson, 2018b). If the true model is a VAR, then a correctly specified VAR is more effi-

[1] As noted in Stock and Watson (2018), LP are direct multistep forecasts. However, the goal of direct multistep forecast is an optimal multistep ahead forecast, whereas the goal of LP is a consistent estimate of the corresponding impulse responses.

[2] In the case of stationary time series, Plagborg-Møller and Wolf (2019) show if the sample size is infinite, linear time-invariant VAR(∞) and LP(∞) estimate the same impulse responses. This equivalence does not hold if the models are augmented with non-linear terms.

cient than LP because VARs impose more structure than LP (Ramey, 2016).[3] If the true model is not a VAR or if the lag length of the VAR is not sufficiently long, then LP can outperform VARs (Plagborg-Møller and Wolf, 2019). Being that LP impulse responses nest VAR impulse responses, the choice of whether to use impulse responses from LP or VARs can be thought of as the bias-variance tradeoff problem with VARs and LP lying on a spectrum of small sample bias variance choices.

It is well known that LP residuals are autocorrelated. Practitioners exclusively estimate LP via OLS with Newey-West standard errors (or some type of Heteroskedastic and Autocorrelation Consistent (HAC) standard errors) (Ramey, 2016). Jordà (2005) argues that since the true data generating process is unknown, Generalized Least Squares (GLS) is not possible and HAC standard errors must be used. Lazarus et al. (2018) claim that LP have to be estimated with HAC standard errors because GLS estimates would be inconsistent.[4] I show that under standard time series assumptions, the autocorrelation process is known and autocorrelation can be corrected for using GLS. Moreover, I show the consistency and asymptotically normality of the LP GLS estimator, as well as the asymptotic efficiency of LP GLS relative to LP OLS.

Being able to specify the autocorrelation process for LP has 3 major implications. First, LP GLS can be substantially more efficient than LP estimated via OLS with Newey-West standard errors. Moreover, once autocorrelation is corrected for, it can be shown that if the data is persistent and the true model is a VAR, LP GLS impulse responses can be approximately as efficient as VAR impulse responses. Whether or not LP GLS impulse responses are approximately as efficient depends on the persistence of the system, the horizon, and the dependence structure of the system. All else equal, the more persistent the system,

[3] If one is willing to assume a likelihood function for the model, this is just the Cramer Rao Lower Bound argument.

[4] Lazarus et al. (2018) assume strict exogeneity (which neither LP or VARs satisfy) is necessary for GLS. Even though strict exogeneity is often assumed for GLS, it is not a necessary condition for GLS (see Hamilton (1994); Stock and Watson (2007a) for discussions on the strict exogeneity assumption for GLS).

the more likely LP GLS impulse responses will be approximately as efficient for horizons typically relevant in practice. It follows that the efficiency of the VAR relative to the LP has been overstated in the literature.

Second, since the autocorrelation process is known, LP GLS can be estimated using fully Bayesian methods.[5] Bayesian LP have many advantages such as allowing the researcher to incorporate prior information for impulse responses at each horizon. Prior information can be used to shrink impulse responses at any horizon to prevent overfitting. Economic theory can be incorporated into the prior to inform the shape of the impulse responses (e.g. the impulse response is monotonic or hump shaped) and to discipline the long-run behavior. Priors can be used to shrink parameter estimates when the number of parameters is large relative to the number of observations making it possible to use LP to estimate systems with big data or panel data with large cross sections over relatively short time frames (e.g. the Eurozone). Moreover, methodologies used for Bayesian VARs (i.e. big data, sparsity, and variable selection methods) can now be carried over to LP. Lastly, Bayesian methods do not need to do anything special to take into account unit roots.

Third, since autocorrelation is explicitly modeled, it is now possible to estimate time-varying parameter LP. Time-varying parameter models are useful for several reasons. Researchers are often interested in whether there is parameter instability in regression models. As noted in Granger and Newbold (1977), macro data encountered in practice are unlikely to be stationary. Stock and Watson (1996) and Ang and Bekaert (2002) show many macroeconomic and financial time series exhibit parameter instability. It is also commonplace for regressions with macroeconomic time series to display heteroskedasticity of unknown form (Stock and Watson, 2007a), and in order to do valid inference, the heteroskedasticity must

[5] Miranda-Agrippino and Ricco (2018) introduce a method called Bayesian LP, but the method is not fully Bayesian, because they replace the estimated scale matrix in the inverse-Wishart posterior with the Newey-West variance covariance matrix. Using plug-in estimates for hyper–parameters is well known to cause probability intervals to underrepresent uncertainty (Koop and Korobilis, 2009; Hoff and Wakefield, 2013). Furthermore, I will show why autocorrelation should be explicitly corrected for in LP.

be taken into account. Parameter instability can occur for many reasons such as policy changes, technological evolution, changing economic conditions, etc. If parameter instability is not appropriately taken into account, it can lead to invalid inference, poor out of sample forecasting, and incorrect policy evaluation. Moreover, as shown in Granger (2008), time-varying parameter models can approximate any non-linear model (non-linear in the variables and/or the parameters), which makes them more robust to model misspecification. Bayesian methods are the primary methods used to estimate time-varying parameter models, and since autocorrelation is explicitly corrected for in Bayesian LP, it is straightforward to apply time-varying parameters to LP.[6]

In this paper, I make several contributions. I show that the autocorrelation process of LP is known and that autocorrelation can be corrected for using GLS. Estimating LP with GLS has three major implications: First, LP GLS can be substantially more efficient and less biased than estimation by OLS with Newey-West standard errors. Second, LP GLS can be estimated using fully Bayesian or frequentist methods. Third, it is now possible to estimate time-varying parameter LP. The paper is outlined as follows: section 1.1 discusses potential drawbacks of OLS estimation of LP. Section 1.2 contains the core result showing that the autocorrelation process of LP is known and illustrates why GLS is possible. Section 1.3 discusses the properties of the GLS estimator. Section 1.4 and 1.5 explains how to estimate LP GLS using both frequentist and Bayesian methods, respectively. Section 1.7 discusses the relative efficiency of LP estimated by OLS with Newey-West standard errors vs LP GLS. Section 1.8 contains Monte Carlo evidence of the small sample properties of LP GLS. Section 1.9 discusses issues in regards to nonstationarity. Section 1.10 explains how time-varying parameter LP can be estimated and illustrates a Bayesian procedure to do so. Section 1.11 concludes.

Some notation: $N(\cdot,\cdot)$, $IW(\cdot,\cdot)$, are the normal, and inverse-Wishart distributions, re-

[6]Time-varying parameter LP do not have to implemented using Bayesian methods.

spectively. $T_n(\cdot,\cdot)$ is the T-distribution with n degrees of freedom. $y_{1:T} = \{y_1, ..., y_T\}$. $\xrightarrow{p}$ is converges in probability, and $\xrightarrow{d}$ is converges in distribution.

1.1 LP and Newey-West Standard Errors

To illustrate how LP work, take the simple VAR(1) model

$$y_t = A_1 y_{t-1} + \varepsilon_t, \tag{1.1}$$

where y_t is a demeaned $r \times 1$ vector of endogenous variables and ε_t is an $r \times 1$ vector white noise process and $var(\varepsilon_t) = \Sigma_\varepsilon$.[7] Assume that the eigenvalues of A_1 have moduli less than unity and $A_1 \neq 0$. Iterating forward leads to

$$y_{t+h} = A_1^{h+1} y_{t-1} + A_1^h \varepsilon_t + \ldots + A_1 \varepsilon_{t+h-1} + \varepsilon_{t+h}.$$

To estimate the impulse responses of a VAR, one would estimate A_1 from equation (1) and then use the non-linear delta method, bootstrapping, or Monte Carlo integration to perform inference on the impulse responses: $\{A_1, A_1^2, \ldots, A_1^{h+1}\}$. To estimate impulse responses using LP, one would estimate the impulse responses directly at each horizon with separate regressions

$$y_t = B_1^{(1)} y_{t-1} + e_t^{(0)},$$

$$y_{t+1} = B_1^{(2)} y_{t-1} + e_{t+1}^{(1)},$$

$$\vdots$$

$$y_{t+h} = B_1^{(h+1)} y_{t-1} + e_{t+h}^{(h)},$$

[7] Without loss of generality, y_t is demeaned in order to remove the constant and simplify notation.

where h is the horizon, and when the true data generating process is a VAR(1), $\{B_1^{(1)}, B_1^{(2)}, \ldots, B_1^{(h+1)}\}$ and $\{A_1, A_1^2, \ldots, A_1^{h+1}\}$ are equivalent. Even if the true data generating process is not a VAR(1), $B_1^{(1)} = A_1$ because the horizon 0 LP is a VAR. In practice, it is common for more than one lag to be used. A VAR(k) and the horizon h LP(k) can be expressed as

$$y_t = A_1 y_{t-1} + \ldots + A_k y_{t-k} + \varepsilon_t,$$

and

$$y_{t+h} = B_1^{(h+1)} y_{t-1} + \ldots + B_k^{(h+1)} y_{t-k} + e_{t+h}^{(h)},$$

respectively. Bear in mind that any VAR(k) can be written as a VAR(1) (companion form), so results and examples involving the VAR(1) the can be generalized to higher order VARs.

LP have been advocated by Jordà (2005) as an alternative to VARs. There are several advantages of using LP as opposed to VARs. First, LP do not constrain the shape of the impulse response function like VARs, so it can be less sensitive to model misspecification (i.e. such as insufficient lag length) because misspecifications are not compounded in the impulse responses when iterating forward.[8] Second, LP can be estimated using simple linear regressions. Third, joint or point-wise analytic inference is simple. Fourth, LP can easily be adapted to handle non-linearities (in the variables or parameters).

LP do have a couple of drawbacks. First, because the dependent variable is a lead, a total of h observations are lost from the original sample when estimating projections for horizon h. Second, the error terms in LP for horizons greater than 0 are inherently autocorrelated. Assuming the true model is a VAR(1), it is obvious that autocorrelation occurs because the

[8] In the case of the linear time-invariant estimators, VAR(∞) and LP(∞) estimate the same impulse responses asymptotically (Plagborg-Møller and Wolf, 2019). This result does not hold if the models are augmented with nonlinear terms.

LP residuals follow an VMA(h) process of the residuals in equation (1). That is

$$e_{t+h}^{(h)} = A_1^h \varepsilon_t + \ldots + A_1 \varepsilon_{t+h-1} + \varepsilon_{t+h},$$

or written in terms of LP

$$e_{t+h}^{(h)} = B_1^{(h)} \varepsilon_t + \ldots + B_1^{(1)} \varepsilon_{t+h-1} + \varepsilon_{t+h}.$$

Frequentists account for the inherent autocorrelation using Newey-West standard errors, which will yield asymptotically correct standard errors in the presence of autocorrelation and heteroskedasticity of unknown forms.[9] Autocorrelation can be corrected for explicitly by including $\{\varepsilon_t, \varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ in the conditioning set of the horizon h LP. Obviously $\{\varepsilon_t, \varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ are unobserved and would have to be estimated, but this issue can be ignored for now and is addressed later.

There are two major advantages of correcting for autocorrelation explicitly. The first is that it fixes what I dub the "increasing variance problem". To my knowledge, the increasing variance problem has not been noticed in the literature. If the true model is a VAR(1), then $var(e_{t+h}^{(h)}) = \sum_{i=0}^{h} A_1^{i\prime} \Sigma_\varepsilon A_1^i$, which is increasing in h.[10] Newey-West standard errors are valid in the presence of autocorrelation because they take into account autocorrelation is present when estimating the covariance matrix; they do not, however, eliminate autocorrelation.[11,12]

[9] This is assuming that a large enough lag truncation parameter for the autocorrelation is chosen. There is a major line of research indicating that Newey-West standard errors perform poorly in small samples with persistent data (Müller, 2014).

[10] Since A_1 has moduli less than unity, geometric progression can be used to show that the sum is bounded asymptotically.

[11] This is a major reason why Kilian and Kim (2011) found that LP had excessive average length relative to the bias-adjusted bootstrap VAR interval in their Monte Carlo simulations. I provide Monte Carlo evidence of this in section 1.8.

[12] Macro variables tend to be persistent, so A_1^i will more likely decay slowly leading to the increase in the variance to be pretty persistent as h increases.

To illustrate, let the true model be an AR(1) with

$$y_t = .99y_{t-1} + \varepsilon_t,$$

where $var(\varepsilon_t) = 1$. The $var(e_{t+h}^{(h)}) = \sum_{i=0}^{h} A_1^{i'} \Sigma_\varepsilon A_1^i = \sum_{i=0}^{h} .99^{2i}$. Table 1.1 presents the asymptotic variance of the residuals for different horizons when estimated by OLS with Newey-West standard errors vs LP estimated with GLS.

Horizons	5	10	20	40
LP NW	5.7093	9.9683	17.3036	28.2102
LP GLS	1	1	1	1

Table 1.1: Asymptotic Variance of Residuals for LP Horizons

Even if Newey-West standard errors are used, the increasing variance problem persists. In terms of the MLE and OLS, correcting for autocorrelation explicitly is asymptotically more efficient because $var(\varepsilon_t) \leq var(e_t^{(h)})$, where the equality only binds when $A_1 = 0$.

The second major advantage of correcting for autocorrelation explicitly is that it helps remedy what I dub the "increased small sample bias problem". When LP are estimated with OLS and Newey-West standard errors, the small sample bias from estimating dynamic models increases relative to the model with no autocorrelation. To see why, let us first review the finite sample bias problem with VARs (see (Pope, 1990) for detailed derivations). Assume the true model is a VAR(1). The OLS estimate for the VAR is

$$\hat{A}_1 = A_1 + \sum_{t=2}^{T} \varepsilon_t y_{t-1}' (\sum_{t=2}^{T} y_{t-1} y_{t-1}')^{-1}.$$

This estimate is biased in finite samples because $E(\sum_{t=2}^{T} \varepsilon_t y_{t-1}' (\sum_{t=2}^{T} y_{t-1} y_{t-1}')^{-1}) \neq 0$ because ε_t and $(\sum_{t=2}^{T} y_{t-1} y_{t-1}')^{-1}$ are not independent. The stronger the correlation between ε_t and $(\sum_{t=2}^{T} y_{t-1} y_{t-1}')^{-1}$, the larger the bias. In macroeconomic applications, the bias is typically downward. The bias disappears asymptotically since ε_t would be correlated with

an increasingly smaller share of $(\sum_{t=2}^{T} y_{t-1}y'_{t-1})^{-1}$.

If one were to estimate a LP via OLS with Newey-West standard errors at horizon h, the OLS estimate would be

$$\hat{B}_1^{(h+1)} = B_1^{(h+1)} + \sum_{t=2}^{T-h} e_{t+h}^{(h)} y'_{t-1} (\sum_{t=2}^{T-h} y_{t-1}y'_{t-1})^{-1}.$$

If one were to correct for autocorrelation by including $\{\varepsilon_t, \varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$, the estimate would be

$$\hat{B}_1^{(h+1)} = B_1^{(h+1)} + \sum_{t=2}^{T-h} \varepsilon_{t+h} y'_{t-1} (\sum_{t=2}^{T-h} y_{t-1}y'_{t-1})^{-1}.$$

The absolute value of the correlation between $e_{t+h}^{(h)}$ and $(\sum_{t=2}^{T-h} y_{t-1}y'_{t-1})^{-1}$ is larger than the absolute value of the correlation between ε_{t+h} and $(\sum_{t=2}^{T-h} y_{t-1}y'_{t-1})^{-1}$ because $e_{t+h}^{(h)} = A_1^h \varepsilon_t + \ldots + A_1 \varepsilon_{t+h-1} + \varepsilon_{t+h}$ is correlated with a larger share of $(\sum_{t=2}^{T-h} y_{t-1}y'_{t-1})^{-1}$.[13] To illustrate, I conduct a simple Monte Carlo simulation where I generate 1,000 samples of length 200 for the following AR(1)

$$y_t = .99 y_{t-1} + \varepsilon_t,$$

where $var(\varepsilon_t) = 1$. I then estimate the impulse responses using a VAR, LP estimated with OLS, and LP estimated with GLS. To correct for autocorrelation using GLS, I include the estimated residuals. Table 1.2 presents the mean impulse responses at different horizons for the different methods.

All of the estimated can be substantially biased, but not correcting for autocorrelation can make the bias substantially worse. Even if autocorrelation is corrected for in LP, there can still be a small sample bias due to the correlation between ε_{t+h} and $(\sum_{t=2}^{T-h} y_{t-1}y'_{t-1})^{-1}$ not be-

[13] This is probably a major reason why Kilian and Kim (2011) found that LP impulse responses were more biased than the VAR impulse responses in their Monte Carlo simulations.

Horizons	5	10	20	40
True	.951	.9044	.8179	.6690
VAR	.8355	.7072	.5231	.3148
LP NW	.8259	.6713	.4223	.0787
LP GLS	.8347	.7045	.5160	.2965

Table 1.2: Mean Impulse Response Estimates for T=200

ing 0 in finite samples, but additional bias due to not explicitly correcting for autocorrelation would be eliminated.[14]

1.2 The Autocorrelation Process of LP

This subsection presents the core result: the autocorrelation process of LP is known under standard time series assumptions and can be corrected for via GLS. First, I will show that even when the true data generating process is not a VAR, including the horizon 0 LP residuals (or equivalently, VAR residuals), $\{\varepsilon_t, \varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$, in the horizon h conditioning set will eliminate autocorrelation as long as the data are stationary and the horizon 0 LP residuals are uncorrelated. Second, I will show that the autocorrelation process of $e_{t+h}^{(h)}$ is known.

Assumption 1. *The data $\{y_t\}$ are stationary and purely non-deterministic so there exists a Wold representation*

$$y_t = \varepsilon_t + \sum_{i=1}^{\infty} \Theta_i \varepsilon_{t-i}.$$

Assumption 1 implies that by the Wold representation theorem, there exists a linear and time-invariant Vector Moving Average (VMA) representation of the uncorrelated one-step ahead forecast errors $\{\varepsilon_t\}$. It follows from the Wold representation theorem that $\varepsilon_t = y_t - Proj(y_t|y_{t-1}, y_{t-2}, \ldots)$ where $Proj(y_t|y_{t-1}, y_{t-2}, \ldots)$ is the (population) orthogonal projection

[14]LP GLS tends to be a little more biased than the VAR because LP estimated at horizon h loses h observations at the end of the sample.

of y_t onto $\{y_{t-1}, y_{t-2}, \ldots\}$.

Consider for each horizon $h = 0, 1, 2, \ldots$ the infinite lag LP

$$y_{t+h} = B_1^{(h+1)} y_{t-1} + B_2^{(h+1)} y_{t-2} + \ldots + e_{t+h}^{(h)}.$$

Proposition 1. *Under Assumption 1, including $\{\varepsilon_t, \varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ in the conditioning set of the horizon h LP will eliminate autocorrelation in the horizon h LP residuals.*

Proof. I first show that

$$Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t, y_{t-1}, y_{t-2}, \ldots) = Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t, y_{t+h-1}, y_{t+h-2}, \ldots).$$

From the Wold representation we know that $\varepsilon_{t+h-1} = y_{t+h-1} - Proj(y_{t+h-1}|y_{t+h-2}, y_{t+h-3}, \ldots)$, which implies that $\{\varepsilon_{t+h-1}, y_{t+h-1}, y_{t+h-2}, y_{t+h-3}, \ldots\}$ are linearly dependent. This implies that y_{t+h-1} can be dropped from $Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t, y_{t+h-1}, y_{t+h-2}, \ldots)$ since it contains redundant information. Therefore,

$$Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t, y_{t+h-1}, y_{t+h-2}, \ldots) = Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t, y_{t+h-2}, y_{t+h-3}, \ldots).$$

Similarly, $\varepsilon_{t+h-2} = y_{t+h-2} - Proj(y_{t+h-2}|y_{t+h-3}, y_{t+h-4}, \ldots)$, which implies that $\{\varepsilon_{t+h-2}, y_{t+h-2}, y_{t+h-3}, y_{t+h-4}, \ldots\}$ are linearly dependent. This implies that y_{t+h-2} can be dropped from $Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t, y_{t+h-2}, y_{t+h-3}, \ldots)$ since it contains redundant information. Therefore,

$$Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t, y_{t+h-2}, y_{t+h-3}, \ldots) = Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t, y_{t+h-3}, y_{t+h-4}, \ldots).$$

This process is repeated until y_t is being dropped due to linear dependence yielding

$$Proj(y_{t+h}|\varepsilon_{t+h-1},\ldots,\varepsilon_t,y_t,y_{t-1},\ldots) = Proj(y_{t+h}|\varepsilon_{t+h-1},\ldots,\varepsilon_t,y_{t-1},y_{t-2},\ldots).$$

Therefore, if the data are stationary and the horizon 0 LP residuals are uncorrelated,

$$Proj(y_{t+h}|\varepsilon_{t+h-1},\ldots,\varepsilon_t,y_{t-1},y_{t-2},\ldots) = Proj(y_{t+h}|\varepsilon_{t+h-1},\ldots,\varepsilon_t,y_{t+h-1},y_{t+h-2},\ldots).$$

Since conditional independence is satisfied it follows that

$$[y_{t+h} - Proj(y_{t+h}|\varepsilon_{t+h-1},\ldots,\varepsilon_t,y_{t-1},y_{t-2},\ldots)] \perp$$

$$[y_{t+h-i} - Proj(y_{t+h-i}|\varepsilon_{t+h-i-1},\ldots,\varepsilon_{t-i},y_{t-i-1},y_{t-i-2},\ldots)]\ \forall i \geq 1,$$

where $\perp$ is the orthogonal symbol. □

Therefore, if the data are stationary and the residuals $\{\varepsilon_t\}$ are uncorrelated, autocorrelation can be eliminated in the horizon h LP by including $\{\varepsilon_t,\varepsilon_{t+1},\ldots,\varepsilon_{t+h-1}\}$ in the conditioning set. Of course, if the true model requires only finitely many lags in the LP specification, then the proof above applies to that case as well, since the longer lags will all have coefficients of zero in population.

Theorem 1. ***The autocorrelation process of the horizon h LP residuals ($e_{t+h}^{(h)}$) is known.***

Proof. We know from the Wold representation that $\varepsilon_t \perp y_{t-1}, y_{t-2},\ldots$, hence $\varepsilon_t \perp \varepsilon_s\, for\, t \neq s$. Recall that the infinite lag horizon h LP is

$$y_{t+h} = B_1^{(h+1)}y_{t-1} + B_2^{(h+1)}y_{t-2} + \ldots + e_{t+h}^{(h)} = Proj(y_{t+h}|y_{t-1},y_{t-2},\ldots) + e_{t+h}^{(h)}. \quad (1.2)$$

By Proposition 1, including $\{\varepsilon_t, \varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ in the conditioning set eliminates autocorrelation, so the horizon h LP can be rewritten as

$$y_{t+h} = Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t, y_{t-1}, y_{t-2}, \ldots) + u_{t+h}^{(h)}, \tag{1.3}$$

where $u_{t+h}^{(h)} = e_{t+h}^{(h)} - Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t) = e_{t+h}^{(h)} - Proj(y_{t+h}|\varepsilon_{t+h-1}) - \ldots - Proj(y_{t+h}|\varepsilon_t)$. The $Proj$ can be broken up additively because $\{\varepsilon_t, \ldots, \varepsilon_{t+h-1}\}$ are orthogonal to each other and to $\{y_{t-1}, y_{t-2}, \ldots\}$. By Proposition 1, $u_{t+h}^{(h)}$ is not autocorrelated. By the Wold representation we know that

$$Proj(y_{t+h}|\varepsilon_t) = \Theta_h \varepsilon_t. \tag{1.4}$$

This implies, the horizon h LP can be written as

$$y_{t+h} = B_1^{(h+1)} y_{t-1} + B_2^{(h+1)} y_{t-2} + \ldots + \Theta_h \varepsilon_t + \ldots + \Theta_1 \varepsilon_{t+h-1} + u_{t+h}^{(h)}, \tag{1.5}$$

which implies

$$e_{t+h}^{(h)} = \Theta_h \varepsilon_t + \ldots + \Theta_1 \varepsilon_{t+h-1} + u_{t+h}^{(h)}.$$

As a result, the autocorrelation process of $e_{t+h}^{(h)}$ is known. Using the same linear dependence arguments as in Proposition 1, it can be shown that

$$Proj(y_{t+h}|\varepsilon_{t+h-1}, \ldots, \varepsilon_t, y_{t-1}, y_{t-2}, \ldots) = Proj(y_{t+h}|y_{t+h-1}, y_{t+h-2}, \ldots),$$

which implies that

$$u_{t+h}^{(h)} = \varepsilon_{t+h},$$

in population. $\square$

Thus in population, the error process is a VMA(h) even if the true model is not a VAR. In population

$$B_1^{(h)} = \Theta_h,$$

which implies

$$e_{t+h}^{(h)} = B_1^{(h)}\varepsilon_t + \ldots + B_1^{(1)}\varepsilon_{t+h-1} + \varepsilon_{t+h}.$$

1.3 LP GLS and Its Properties

Since $e_{t+h}^{(h)}$ can be written as

$$e_{t+h}^{(h)} = B_1^{(h)}\varepsilon_t + \ldots + B_1^{(1)}\varepsilon_{t+h-1} + u_{t+h}^{(h)}, \tag{1.6}$$

GLS can be used to eliminate autocorrelation in LP while avoiding increasing the number of parameters by including $\{\varepsilon_t, \varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ in the horizon h conditioning set. To understand how, I'll first explain what happens when $\{\varepsilon_t, \varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ is included in the conditioning set. Just like it is impossible to estimate a VAR(∞) in practice, one cannot estimate LP with infinite lags since there is insufficient data. In practice truncated LP are used where the lags are truncated at k. The proofs of consistency and asymptotic normality discuss the rate at which k needs to grow with the sample size to ensure consistent estimation of the impulse responses. In practice, k, needs to be large enough that the estimated residuals from the horizon 0 LP are uncorrelated, which is what will be assumed for now.

From Theorem 1 we know the horizon h LP is

$$y_{t+h} = B_1^{(h+1)} y_{t-1} + \ldots + B_k^{(h+1)} y_{t-k} + B_1^{(h)} \varepsilon_t + \ldots + B_1^{(1)} \varepsilon_{t+h-1} + u_{t+h}^{(h)}. \tag{1.7}$$

Due to $\{\varepsilon_t, \varepsilon_{t+1}, \ldots, \varepsilon_{t+h-1}\}$ being unobserved, the estimates $\{\hat{\varepsilon}_t, \hat{\varepsilon}_{t+1}, \ldots, \hat{\varepsilon}_{t+h-1}\}$ from the horizon 0 LP must be used instead. Estimates of the impulse responses are still consistent (see appendix for proof), however, even if the sample size is large, inference on the parameters will underrepresent uncertainty because $\{\hat{\varepsilon}_t, \hat{\varepsilon}_{t+1}, \ldots, \hat{\varepsilon}_{t+h-1}\}$, are generated regressors (Pagan, 1984). In order to do valid inference, one must take into account that the generated regressors were estimated.[15]

For now, I will ignore the additional uncertainty from the generated regressors $\{\hat{\varepsilon}_t, \hat{\varepsilon}_{t+1}, \ldots, \hat{\varepsilon}_{t+h-1}\}$. Including $\{\hat{\varepsilon}_t, \hat{\varepsilon}_{t+1}, \ldots, \hat{\varepsilon}_{t+h-1}\}$ in the conditioning set increases the number of parameters in each equation in the system by $h \times r$. If consistent estimates of $\{\hat{B}_1^{(h)}, \hat{B}_1^{(h-1)}, \ldots, \hat{B}_1^{(1)}\}$ are obtained in previous horizons, one can do a Feasible GLS (FGLS) transformation. Let $\widetilde{y}_{t+h}^{(h)} = y_{t+h} - \hat{B}_1^{(h)} \hat{\varepsilon}_t - \ldots - \hat{B}_1^{(1)} \hat{\varepsilon}_{t+h-1}$. Then one can estimate horizon h via the following equation

$$\widetilde{y}_{t+h}^{(h)} = B_1^{(h+1)} y_{t-1} + \ldots + B_k^{(h+1)} y_{t-k} + \tilde{u}_{t+h}^{(h)}. \tag{1.8}$$

$\widetilde{y}_{t+h}^{(h)}$ is just a GLS transformation that eliminates the autocorrelation problem in LP without having to sacrifice degrees of freedom and $\tilde{u}_{t+h}^{(h)}$ is the error term corresponding to this GLS transformation. If the impulse responses are estimated consistently, then by the continuous mapping theorem, $\widetilde{y}_{t+h}^{(h)}$ converges in probability to the true GLS transformation $y_{t+h}^{(h)} = y_{t+h} - B_1^{(h)} \varepsilon_t - \ldots - B_1^{(1)} \varepsilon_{t+h-1}$ asymptotically. For clarification LP can be estimated sequentially

[15] In the proof of asymptotic normality of the limiting distribution, it can be seen that the impact of the generated regressors does not disappear asymptotically.

horizon by horizon as follows. First estimate the horizon 0 LP

$$y_t = B_1^{(1)} y_{t-1} + \ldots + B_k^{(1)} y_{t-k} + u_t^{(0)},$$

and due to the horizon 0 LP being a VAR $\varepsilon_t = u_t^{(0)}$. $\hat{B}_1^{(1)}$ and $\hat{\varepsilon}_t$ are estimates of $B_1^{(1)}$ and ε_t respectively . Horizon 1 can be estimated as

$$\widetilde{y}_{t+1}^{(1)} = B_1^{(2)} y_{t-1} + \ldots + B_k^{(2)} y_{t-k} + \tilde{u}_{t+1}^{(1)},$$

where $\tilde{y}_{t+1}^{(1)} = y_{t+1} - \hat{B}_1^{(1)} \hat{\varepsilon}_t$, and $\hat{B}_1^{(2)}$ is the GLS estimate of $B_1^{(2)}$. Horizon 2 can be estimated as

$$\tilde{y}_{t+2}^{(2)} = B_1^{(3)} y_{t-1} + \ldots + B_k^{(3)} y_{t-k} + \tilde{u}_{t+2}^{(2)},$$

where $\widetilde{y}_{t+2}^{(2)} = y_{t+2} - \hat{B}_1^{(2)} \hat{\varepsilon}_t - \hat{B}_1^{(1)} \hat{\varepsilon}_{t+1}$, and $\hat{B}_1^{(3)}$ is the GLS estimate of $B_1^{(3)}$. Horizon 3 can be estimated as

$$\tilde{y}_{t+3}^{(3)} = B_1^{(4)} y_{t-1} + \ldots + B_k^{(4)} y_{t-k} + \tilde{u}_{t+3}^{(3)},$$

where $\widetilde{y}_{t+3}^{(3)} = y_{t+3} - \hat{B}_1^{(3)} \hat{\varepsilon}_t - \hat{B}_1^{(2)} \hat{\varepsilon}_{t+1} - \hat{B}_1^{(1)} \hat{\varepsilon}_{t+2}$, and so on.

The LP GLS estimator has the following three properties:

Theorem 2. *Under the assumptions stated in Lewis and Reinsel (1985), the LP GLS estimator is consistent. In particular*

$$\hat{B}_1^{(h)} \xrightarrow{p} \Theta_h.$$

Theorem 3. *Under the assumptions stated in Lewis and Reinsel (1985), the limiting distribution of the LP GLS estimates are asymptotically normal.*

Theorem 4. *Under the assumptions stated in Lewis and Reinsel (1985), the limiting distribution of the LP GLS estimates are asymptotically more efficient than the limiting distribution of the LP OLS estimates.*

Under the assumptions stated in Lewis and Reinsel (1985), Jordà and Kozicki (2011) show the consistency and asymptotically normality of $\{B_1^{(h+1)}, B_2^{(h+1)}, \ldots, B_k^{(h+1)}\}$ when estimated via OLS.

Remark. The assumptions are general enough to include most stationary invertible VARMA models. Jordà and Kozicki (2011) proof is an extension of Lewis and Reinsel (1985), who show consistency and asymptotic normality of the VAR(∞). The conditions in Lewis and Reinsel (1985) state the rate at which the lag length, k, needs to grow in order for the estimates to be consistent and asymptotically normal.

Proof. See appendix. The explicit assumptions and proofs are in section A.4 of the appendix. □

As noted earlier, the parameters used in the GLS correction are not known, and their uncertainty should be taken into account in order to do valid inference. To take into account the uncertainty in the generated regressors, frequentist can use bootstrapping, multi-step estimation (Murphy and Topel, 1985), or joint estimation (Newey and McFadden, 1994). Bayesian's can marginalize uncertainty via Monte Carlo integration. Estimation for both frequentist and Bayesian methods will be discussed in the next 2 sections.

1.4 Frequentist Estimation via Bootstrapping

For frequentist estimation, LP GLS can be implemented using a circular block bootstrap scheme (Politis and Romano, 1994). Bootstrap samples are first created using the circu-

lar block scheme, then for each bootstrap sample, FGLS estimates of the LP horizons are constructed. To illustrate, first one must decide on the number of bootstrap draws, J, the maximum number of of impulse response horizons to be estimated, H, and the number of consecutive blocks, L. There are no good rules of thumb for choosing L in general, so I follow Berkowitz et al. (1999) and set $L = T^{1/3}$. To construct the bootstrap data sets, the original data, $\{y_1, y_2, y_{t-2}, \ldots, y_T\}$, is wrapped around in a circle so that y_1 follows y_T. By construction, the horizon h LP depends on the $\{y_{t+h}, y_{t-1}, y_{t-2}, \ldots, y_{t-k}\}$ tuple. Since LP will be estimated using FGLS and the transformation must be done using the same data, one must first construct all possible $\{y_{t+H}, \ldots, y_{t-k}\}$ tuples. Then to preserve correlation in the data blocks of L consecutive tuples are drawn at random and concatenated to generate a bootstrap sample. Then for each bootstrap sample, the impulse responses $\{B_1^{(h+1)}, B_2^{(h+1)}, \ldots, B_k^{(h+1)}\}$ are estimated for each horizon using the FGLS estimation described in the previous section. This is done for each of the J bootstrap draws. To clarify,

Algorithm 1: Block Bootstrapping Without Bias Adjustment

1: for each bootstrap replication $j = 1, \ldots, J$
2: draw blocks of L consecutive $\{y_{t+H}, y_{t-1}, y_{t-2}, \ldots, y_{t-k}\}$ tuples to generate a bootstrap sample.
3: estimate $\{B_1^{(h+1)}, B_2^{(h+1)}, \ldots, B_k^{(h+1)}\}$ for each horizon via the FGLS procedure.
4: end.

Denoting $\{B_1^{(h+1),<j>}, B_2^{(h+1),<j>}, \ldots, B_k^{(h+1),<j>}\}$ as jth bootstrap replication for the impulse responses, 95% confidence intervals can then be constructed by taking the 2.5% and 97.5% quantiles of the parameter(s) of interest. The bootstrap can also be implemented with bias adjustment. The bias adjustment of the LP GLS bootstrap follows the general procedure of Efron and Tibshirani (1993), and bias adjustment is implemented for each horizon. The bias of any parameter, for example $B_1^{(h)}$, can be calculated via $bias = J^{-1}(\sum_{j=1}^{J} B_1^{(h),<j>}) - \hat{B}_1^{(h)}$. The bias adjusted bootstrap replications are then $B_1^{(h),<j>,BA} = B_1^{(h),<j>} - bias$. To

summarize how to block bootstrap with bias adjustment,

Algorithm 2: Block Bootstrapping With Bias Adjustment

```
for each bootstrap replication j = 1, ..., J
    draw blocks of L consecutive {y_{t+H}, ..., y_{t-k}} tuples to generate a bootstrap sample.
end
for each LP horizon h = 0, ..., H - 1
    for each bootstrap replication j = 1, ..., J
        estimate {B_1^{(h+1)}, B_2^{(h+1)}, ..., B_k^{(h+1)}} via the FGLS procedure.
    end
    calculate the bias and bias adjust the bootstrap estimates as in
        Efron and Tibshirani (1993).
end
```

The Monte Carlo simulations analyzing the finite sample properties implements the bias adjustment version of the bootstrap.

1.5 Bayesian Estimation

1.5.1 The Likelihood

Despite LP growing in popularity, there has not been a fully Bayesian treatment, which is probably due to the belief that Newey-West standard errors must be used because the autocorrelation process is unknown. Since LP can be estimated using GLS, it is now possible to give a fully Bayesian treatment of LP. Due to LP being standard linear regressions, one just needs to be able to set up the likelihood and elicit a prior. The default prior used in this

paper is the conjugate normal inverse-Wishart prior. Conjugate priors need not be used. LP are linear regressions, so any prior that can be used with a linear regression can be used with Bayesian LP.

LP at horizon 0 are equivalent to VARs. To estimate LPs at horizon 0 just estimate

$$y_t = B_1^{(1)} y_{t-1} + B_2^{(1)} y_{t-2} + \ldots + B_k^{(1)} y_{t-k} + u_t^{(0)},$$

as one would a standard Bayesian VAR. Define $\beta^{(0)} \equiv vec(\left[B_1^{(1)}, B_2^{(1)}, \ldots, B_k^{(1)}\right]')$, $X_t^{'(0)} \equiv I_n \otimes [y_{t-1}, y_{t-2}, \ldots, y_{t-k}]'$, then

$$y_t = X_t^{'(0)} \beta^{(0)} + u_t^{(0)},$$

where $u_t^{(0)} \sim N(0, \Sigma_u^{(0)})$. Assume a conditional normal inverse-Wishart prior for $p(\beta^{(0)}, \Sigma_u^{(0)})$. That is

$$p(\beta^{(0)} | \Sigma_u^{(0)}) \sim N(\underline{b}, \Sigma_u^{(0)} \otimes \underline{\Omega}),$$

$$p(\Sigma_u^{(0)}) \sim IW(\underline{n}, \underline{\Psi}),$$

where $\underline{b}$, $\underline{\Omega}$, $\underline{\Psi}$, and $\underline{n}$ are prior hyperparameters. The posterior is also conditional normal inverse-Wishart

$$p(\beta^{(0)} | \Sigma_u^{(0)}, y_{1:T}) \sim N(\overline{b}, \Sigma_u^{(0)} \otimes \overline{\Omega}),$$

$$p(\Sigma_u^{(0)} | y_{1:T}) \sim IW(\overline{n}, \overline{\Psi}),$$

where $\overline{b}$, $\overline{\Omega}$, $\overline{\Psi}$, and $\overline{n}$ are posterior hyperparameters whose formulas are well known and can be found in the appendix. After estimating horizon 0, one can obtain J posterior draws of residuals $\{\varepsilon_{k+1}, \ldots, \varepsilon_T\}$ using the fact that $\varepsilon_t^{<j>} = y_t - X_t^{'(0)} \beta^{(0),<j>}$, where $\beta^{(0),<j>}$ is the

jth posterior draw of $\beta^{(0)}$. Now posterior draws of $y_{t+1}^{(1)}$ can be constructed via $\tilde{y}_{t+1}^{(1),<j>} = y_{t+1} - B_1^{(1),<j>}\varepsilon_t^{<j>}$. To understand why posterior draws for $y_{t+1}^{(1)}$ are needed, note that in GLS, one uses parameter estimates in the transformation and treat the transformation as known. The transformation, however, does not take into account uncertainty in the parameters used for the transformation, so to properly take into account uncertainty, one must marginalize out uncertainty in the transformation.

For each J, define $y_{t+1}^{(1)} \equiv \tilde{y}_{t+1}^{(1),<j>}$ and $B_1^{(1)} \equiv B_1^{(1),<j>}$, which means for each J we treat the GLS transformation as known. The horizon 1 LP is

$$y_{t+1}^{(1)} = B_1^{(2)} y_{t-1} + B_2^{(2)} y_{t-2} + \ldots + B_k^{(2)} y_{t-k} + u_{t+1}^{(1)},$$

where $u_{t+1}^{(1)} \sim N(0, \Sigma_u^{(1)})$. Define $\beta^{(1)} \equiv vec(\left[B_1^{(2)}, B_2^{(2)}, \ldots, B_k^{(2)}\right]')$ and $X_t^{'(1)} \equiv I_n \otimes [y_{t-1}, y_{t-2}, \ldots, y_{t-k}]'$. Then the horizon 1 LP can be rewritten as

$$y_{t+1}^{(1)} = X_t^{'(1)} \beta^{(1)} + u_{t+1}^{(1)}.$$

Again, assume a conditional normal inverse Wishart prior for $p(\beta^{(1)}, \Sigma_u^{(1)})$

$$p(\beta^{(1)} | \Sigma_u^{(1)}) \sim N(\underline{b}^{(1)}, \Sigma_u^{(1)} \otimes \underline{\Omega}^{(1)}),$$

$$p(\Sigma_u^{(1)}) \sim IW(\underline{n}^{(1)}, \underline{\Psi}^{(1)}).$$

The posterior is conditional normal inverse-Wishart. That is

$$p(\beta^{(1)} | \Sigma_u^{(1)}, y_{1:T}) \sim N(\overline{b}^{(1)}, \Sigma_u^{(1)} \otimes \overline{\Omega}^{(1)}),$$

$$p(\Sigma_u^{(1)} | y_{1:T}) \sim IW(\overline{n}^{(1)}, \overline{\Psi}^{(1)}).$$

One Monte Carlo draw is obtained from the conditional posterior for each J, which marginal-

izes out uncertainty in the GLS transformation.

This is done at each horizon in the LP. Before estimation of horizon h, one can obtain posterior draws of $y_{t+h}^{(h)}$ via $\tilde{y}_{t+h}^{(h),<j>} = y_{t+h} - B_1^{(h),<j>}\varepsilon_t^{<j>} - \ldots - B_1^{(1),<j>}\varepsilon_{t+h-1}^{<j>}$. For each J, define $y_{t+h}^{(h)} \equiv \tilde{y}_{t+h}^{(h),<j>}$ and $B_1^{(1)} \equiv B_1^{(1),<j>}, \ldots, B_1^{(h)} \equiv B_1^{(h),<j>}$. The horizon h LP is

$$y_{t+h}^{(h)} = B_1^{(h+1)} y_{t-1} + B_2^{(h+1)} y_{t-2} + \ldots + B_k^{(h+1)} y_{t-k} + u_{t+h}^{(h)},$$

where $u_{t+h}^{(h)} \sim N(0, \Sigma_u^{(h)})$. Define $\beta^{(h)} \equiv vec(\left[B_1^{(h+1)}, B_2^{(h+1)}, \ldots, B_k^{(h+1)}\right]')$, $X_t^{'(h)} \equiv I_n \otimes [y_{t-1}, y_{t-2}, \ldots, y_{t-k}]'$. The horizon h LP can be rewritten as

$$y_{t+h}^{(h)} = X_t^{'(h)} \beta^{(h)} + u_{t+h}^{(h)}.$$

Again, assume a conditional normal inverse gamma prior for $p(\beta^{(h)}, \Sigma_u^{(h)})$

$$p(\beta^{(h)} | \Sigma_u^{(h)}) \sim N(\underline{b}^{(h)}, \Sigma_u^{(h)} \otimes \underline{\Omega}^{(h)}),$$

$$p(\Sigma_u^{(h)}) \sim IW(\underline{n}^{(h)}, \underline{\Psi}^{(h)}).$$

The posterior is conditional normal inverse-Wishart

$$p(\beta^{(h)} | \Sigma_u^{(h)}, y_{1:T}) \sim N(\overline{b}^{(h)}, \Sigma_u^{(h)} \otimes \overline{\Omega}^{(h)}),$$

$$p(\Sigma_u^{(h)} | y_{1:T}) \sim IW(\overline{n}^{(h)}, \overline{\Psi}^{(h)}).$$

One Monte Carlo draw is obtained from the conditional posteriors for each J, which marginalized out uncertainty in the GLS transformation. To summarize,

Algorithm 3: Bayesian LP

```
Estimate the Bayesian VAR/horizon 0 LP.
Generate J posterior draws for {B_1^(1), B_2^(1), ..., B_k^(1)}
for each LP horizon h = 1, ..., H − 1
    for each posterior draw j = 1, ..., J
        estimate {B_1^(h+1), B_2^(h+1), ..., B_k^(h+1)} via the Bayesian version of the FGLS
            procedure.
    end
end
```

1.5.2 Priors

Bayesian LP allow the researcher to incorporate prior information for impulse responses at each horizon. Incorporating prior information has multiple advantages. Prior information can be used to shrink impulse responses at any horizon to prevent overfitting, which is often desirable in forecasting or when the number of parameters is large (Giannone et al., 2015). Economic theory can be incorporated into the prior to inform the shape of the impulse responses (e.g. the impulse response is monotonic or hump shaped) and to discipline the long-run behavior (Giannone et al., 2018). Prior information from economic theory can also be used to smooth impulse responses across horizons, which may be desirable in certain contexts (Barnichon and Brownlees, 2018; Stock and Watson, 2018).

The default prior used in this paper is a conjugate training sample prior. When using a training sample prior in Bayesian LP, the researcher must decide how many horizons they are going to estimated before they choose the size of the training sample. To understand why, assume that the training sample is of size $\underline{T}$. The same training sample must be used for each horizon, so the training sample must be large enough to estimate a training sample prior at each horizon. Recall that when estimating horizon h, h observations will be lost

from the original sample, so the training sample for horizon h has $\underline{T} - h$ observations.[16]

As shown in the previous section, the horizon 0 LP

$$y_t = B_1^{(1)} y_{t-1} + B_2^{(1)} y_{t-2} + \ldots + B_k^{(1)} y_{t-k} + u_t^{(0)}.$$

can be recast as

$$y_t = X_t^{'(0)} \beta^{(0)} + u_t^{(0)},$$

where $\beta^{(0)} \equiv vec(\left[B_1^{(1)}, B_2^{(1)}, \ldots, B_k^{(1)}\right]')$, $X_t^{'(0)} \equiv I_n \otimes [y_{t-1}, y_{t-2}, \ldots, y_{t-k}]'$, and $u_t^{(0)} \sim N(0, \Sigma_u^{(0)})$. The conjugate training sample prior for $p(\beta^{(0)}, \Sigma_u^{(0)})$ is

$$p(\beta^{(0)} | \Sigma_u^{(0)}) \sim N(\underline{b}, \Sigma_u^{(0)} \otimes \underline{\Omega}),$$

$$p(\Sigma_u^{(0)}) \sim IW(\underline{n}, \underline{\Psi}).$$

$\underline{n}$ is the prior degrees of freedom, $\underline{b} = \hat{\beta}_{OLS}$ and $\underline{\Psi} = \underline{n}\hat{\Sigma}_{OLS}$, where $\hat{\beta}_{OLS}$ and $\hat{\Sigma}_{OLS}$ are the OLS results from the training sample. $\underline{\Omega} = \frac{\underline{T}}{\underline{n}}(\underline{X}'\underline{X})^{-1}$ where $\underline{X}$ is the design matrix for the training sample and $\frac{\underline{T}}{\underline{n}}$ rescales the conditional variance of $\beta^{(0)}$ so the conditional distribution will have the asymptotic variance of the OLS results based on the average of $\underline{n}$ observations.[17] $\underline{n}$, which determines the informativeness of the prior, can be chosen by the researcher or a prior can be placed on $\underline{n}$ and estimated using Griddy Gibbs or sampling importance resampling. In order for the prior mean of $\Sigma_u^{(0)}$ to be defined, $\underline{n} \geq p + 2$. By default, I set $\underline{n} = p + 2$ to make the prior weakly informative but still proper. The diagonal of $\underline{\Omega}$ can be taken to prevent collinearity issues if the prior is only based on small training sample (Brodersen et al., 2015). When estimating the training sample prior for horizons 1

[16]This does not account for the k presample observations that will be treating as deterministic in the VAR(k).

[17]This is in the spirit of the unit information prior (Kass and Wasserman, 1995), but since this is done over a training sample, it does not make double use of the data.

and greater, autocorrelation is corrected for in the training sample estimates using the GLS procedure discussed in Section 1.3.[18]

Even though the conjugate normal inverse-Wishart training sample prior is the only prior presented, many priors can be used with Bayesian LP. The priors need not be conjugate. LP are linear regressions, so any prior that can be used with a linear regression can be used with Bayesian LP. Again, Bayesian LP allow the researcher to incorporate prior information for impulse responses at each horizon. Prior information can be used to shrink impulse responses at any horizon to prevent overfitting. Economic theory can be incorporated into the prior to inform the shape of the impulse responses and to discipline the long-run behavior, which would help smooth impulse responses across horizons and alleviate the sometimes erratic impulse responses estimated from frequentist LP.

1.6 Structural Identification

This subsection briefly discusses structural identification in LP GLS. These techniques can be applied to both the bootstrapped LP and Bayesian LP. For an extensive review of structural identification in VARs and LP see Ramey (2016), and for an extensive treatment of identification in VARs and LP using external instruments see Stock and Watson (2018). Structural identification in Bayesian LP is essentially the same as identification with frequentist LP. Going back to the horizon 0 LP

$$y_t = B_1^{(1)} y_{t-1} + B_2^{(1)} y_{t-2} + \ldots + B_k^{(1)} y_{t-k} + u_t^{(0)},$$

let $u_t^{(0)} = Rs_t$ where s_t is a vector of structural shocks and R is an invertible matrix. If R is known, after estimating $\{B_1^{(1)}, B_1^{(2)}, \ldots, B_1^{(h+1)}\}$, one can construct the structural

[18] Uncertainty is not marginalized out in the GLS transformation, $\widetilde{y}_{t+h}^{(h)}$, for the training sample.

impulse responses, $\{G^{(1)}, G^{(2)}, \ldots, G^{(h+1)}\}$, via Monte Carlo integration where $G^{(h)} = B_1^{(h)} R$. Typically R is not known, but can be estimated, so Monte Carlo integration can still be applied. An example of R being estimated would be a triangular (recursive) ordering.[19] One would estimated horizon 0 LP, and then apply a recursive ordering to posterior (or bootstrap) draws of $\Sigma_u^{(0)}$ to obtain draws of R, and then draws of $G^{(h)}$ can be constructed via $G^{(h)} = B_1^{(h)} R$.

It is often the case that the researcher may not know all of the identifying restrictions in R or may believe that R is not invertible, but the researcher has an instrument that they believe can trace out impulse responses of interest. The impulse responses of interest can instead be estimated by LP instrumental variable regressions (LP-IV). Stock and Watson (2018) show that in order for LP-IV to be valid, 3 conditions need to be satisfied. Decompose s_t into $s_{1,t}$ and $s_{2,t}$ where $s_{1,t}$ is the structural shock of interest at time t and $s_{2,t}$ represents all other structural shocks at time t. Let z_t be an instrument that the researcher believes can trace out the impulse responses of $s_{1,t}$. The instrument must satisfy the following three conditions

$$E[s_{1,t} z_t] \neq 0,$$

$$E[s_{2,t} z_t] = 0,$$

$$E[s_{t+j} z_t] = 0 \, for \, j \neq 0.$$

The first two conditions are just the standard relevance and exogeneity conditions for instrumental variable regression. The third condition is a lead-lag exogeneity condition, which guarantees that the instrument, z_t, is only identifying the impulse response of the shock $s_{1,t}$. If the third condition is not satisfied, then z_t will amalgamate the impulse responses at different horizons. It may be the case that these conditions are only satisfied after conditioning

[19] In the literature a triangular (recursive) ordering is often called a cholesky ordering because people often apply a cholesky decomposition to impose the ordering. It should be noted that the cholesky normalizes the variances of the structural shocks to unity. If one does not want to normalize the structural shocks one can instead use the LDL decomposition to impose recursive the ordering.

on suitable control variables (e.g. the lags of a VAR/horizon 0 LP).

Frequentist typically estimate LP-IV via two-stage least squares (2SLS). For example, say I want to estimate the impulse response, $g^{(h)}$, the impact a shock to monetary policy has on output at horizon h. Let output be denoted as $output_t$ and the monetary policy variable mp_t. The frequentists approach is to estimate LP-IV by running

$$output_{t+h} = g^{(h)} mp_t + control\, variables + error^{(h)}_{t+h} \tag{1.9}$$

via 2SLS and using z_t as an instrument for mp_t. Newey and West (1987) standard errors are used to account for autocorrelation, but as shown section 1.1, this ignores the increasing variance problem. The increasing variance problem is particularly problematic with LP-IV because the increasing variance can weaken the strength of instrument for $h \geq 1$.[20] Alternatively, the impulse responses of shocks to $s_{1,t}$ can be recovered if z_t is included as an endogenous variable in the system and ordered first (Paul, ming; Plagborg-Møller and Wolf, 2019). Let $\mathring{y}_t = \begin{bmatrix} z_t \\ y_t \end{bmatrix}$ where y_t contains mp_t, $output_t$, and the control variables at time t, then the horizon 0 LP/VAR is

$$\mathring{y}_t = \mathring{B}^{(1)}_1 \mathring{y}_{t-1} + \mathring{B}^{(1)}_2 \mathring{y}_{t-2} + \ldots + \mathring{B}^{(1)}_k \mathring{y}_{t-k} + \mathring{u}^{(0)}_t.$$

Since z_t is ordered first due to its exogeneity, the residual for the z_t equation, $\mathring{u}^{(0)}_{1,t}$, will be able to trace out the structural impulse responses of interest.[21] Going back to the monetary policy example, the impulse response $g^{(h)}$ can be constructed as the ratio of the impulse response of $output_{t+h}$ to $\mathring{u}^{(0)}_{1,t}$ divided the impulse response of mp_t to $\mathring{u}^{(0)}_{1,t}$. Hence by imbedding z_t as

[20] Whether the strength of the instrument is weakened depends in part on the type of impulse response being estimated. For example if one is estimated a cumulative multiplier directly like in Ramey and Zubairy (2018), the autocorrelation would weaken the strength of the instrument since the first stage of the 2SLS procedure has an increasing variance problem.

[21] Even if the control variables are exogenous to the system, any VARX can be written as a VAR with the exogenous variables ordered first in a block recursive scheme, therefore estimates from this setup are consistent.

an endogenous variable in the system and ordering it first, one can just estimate equation (2) via their preferred LP GLS method and construct the impulse responses of interest.

1.7 LP GLS and Relative Efficiency

To give a sense of potential efficiency gains of estimating LP via GLS, I will compare the asymptotic relative efficiency of the LP GLS estimator and the standard LP estimator when the true model is an AR(1). The asymptotic results apply for frequentists and Bayesians estimation due to the Bernstein-Von Mises theorem. Take the simple AR(1) model

$$y_t = ay_{t-1} + \varepsilon_t,$$

where $|a|< 1$ and $a \neq 0$ and ε_t is a white noise error process with $E(\varepsilon_t) = 0$ and $var(\varepsilon_t) = \sigma^2$. This implies that $E(y_t) = 0$ and the $var(y_t) = E(y_t' y_t) = \frac{\sigma^2}{(1-a^2)}$. Define $\{b^{(1)}, b^{(2)}, \ldots, b^{(h+1)}\}$ as the LP impulse responses for the AR(1) model. The limiting distribution of the LP GLS impulse response at horizon h is

$$\sqrt{T}(\hat{b}^{(h)} - a^h) \xrightarrow{d} N(0, [1 + (h^2 - 1)a^{2h-2}](1 - a^2)),$$

(follows from Theorem 4). The limiting distribution of the LP impulse response estimated by OLS with Newey-West standard errors at horizon h is

$$\sqrt{T}(\hat{b}^{(h)} - a^h) \xrightarrow{d} N(0, (1 - a^2)^{-1}[1 + a^2 - \{2h + 1\}a^{2h} + \{2h - 1\}a^{2h+2}]),$$

(Bhansali, 1997). The relative efficiency between the LP GLS and LP impulse responses,

$$\frac{[1 + (h^2 - 1)a^{2h-2}](1 - a^2)^2}{[1 + a^2 - \{2h + 1\}a^{2h} + \{2h - 1\}a^{2h+2}]},$$

Horizons	3	5	10	20	30	40
$a = .99$	.993	.979	.945	.88	.818	.759
$a = .975$	.983	.948	.864	.713	.580	.464
$a = .95$	.966	.896	.735	.475	.288	.165
$a = .9$	.931	.792	.508	.179	.061	.029
$a = .75$	.827	.53	.195	.123	.123	.123
$a = .5$	.727	.496	.45	.45	.45	.45
$a = .25$	.854	.828	.827	.827	.827	.827
$a = .1$	.971	.97	.97	.97	.97	.97
$a = .01$	1	1	1	1	1	1

Table 1.3: Relative Efficiency of LP (GLS) to LP (NW)

determines which specification is more efficient. Note that the relative efficiency not only depends on the persistence, a, but on the horizon as well.

The gains from LP GLS can be large but they are not necessarily monotonic. This is because if the persistence is not that high, the impulse responses decay to zero quickly making the variance of the impulse responses small, and the gains from correcting for correcting for autocorrelation are not as large.

The efficiency gains of estimating LP via GLS, do not stop there. It turns out that when the true model is a AR(1) and the system is persistent enough, LP estimated with GLS can be approximately as efficient as the AR(1). Let $\hat{a}$ be the OLS estimate, the OLS estimate of a has the limiting distribution

$$\sqrt{T}(\hat{a} - a) \xrightarrow{d} N(0, 1 - a^2).$$

By the delta method, the horizon h impulse response has the limiting distribution

$$\sqrt{T}(\hat{a}^h - a^h) \xrightarrow{d} N(0, h^2 a^{2h-2}(1 - a^2)).$$

Horizons	5	10	20	30	40
$a = .99$	.997	.998	.999	.999	.999
$a = .975$	.991	.994	.996	.996	.996
$a = .95$	.980	.985	.985	.980	.968
$a = .9$	.95	.946	.881	.667	.302
$a = .75$	.736	.362	.007	0	0
$a = .5$	0	0	0	0	0

Table 1.4: Asymptotic Relative Efficiency of AR to LP (GLS)

The asymptotic relative efficiency between the AR and LP GLS impulse responses

$$\frac{h^2 a^{2h-2}}{h^2 a^{2h-2} + (1 - a^{2h-2})},$$

determines which specification is more efficient. Since the true model is an AR(1), if the errors are normal, the AR(1) model will be asymptotically more efficient due to the Cramer-Rao lower bound (Bhansali, 1997). Table 1.4 of the relative efficiency between the AR and LP impulse responses for different values of a.

If the data is persistent enough, the LP impulse responses have approximately the same variance for horizons relevant in macro. For example, the economics profession has still not determined if GDP has a unit root or not. Assume that GDP is stationary but highly persistent with an AR(1) coefficient of .99. In this case, the AR(1) impulse responses has approximately the same variance for at least the first 40 horizons. Müller (2014) estimates the AR(1) coefficient for unemployment to be approximately .973. This would lead to the AR(1) impulse responses having approximately the same variance for at least the first 40 horizons. Other important macroeconomic variables such as inflation and the 3 month interest rate and most macro aggregates are also highly persistent and would display similar results. It is not until the AR(1) coefficient is .9 that you can see a notable difference over the first 40 horizons, and even then it is not until about 20 or so horizons out.

When the true model is a multivariate VAR things become more complicated. Efficiency

still depends on the horizon and persistence, but because persistence can vary across the equations in the system, then for any horizon, LP could be approximately as efficient for some impulse responses and much less efficient for others. To see why, let us return to the VAR(1) model

$$y_t = A_1 y_{t-1} + \varepsilon_t.$$

Take the eigenvalue decomposition of $A_1 = E\Lambda_1 E'$, where $\Lambda_1 = diag(\lambda_1, \ldots, \lambda_k)$ is the diagonal matrix of distinct nonzero eigenvalues and $E = [e_1, \ldots, e_k]$ is the corresponding eigenmatrix and $EE^{-1} = I$ where I is the identity matrix. As a result $A_1^h = E\Lambda_1^h E' = \sum \lambda_1^h e_1 e_1'$. Define $w_t = E^{-1} y_t$ and $\eta_t = E^{-1}\varepsilon_t$. For simplicity assume E is known. This implies the VAR can be transformed into

$$w_t = \Lambda_1 w_{t-1} + \eta_t,$$

which will be called the transformed model. Consequently

$$w_{t+h} = \Lambda_1^{h+1} w_{t-1} + \Lambda_1^h \eta_t + \ldots + \Lambda_1 \eta_{t+h-1} + \eta_{t+h}.$$

Since Λ_1 is diagonal, each equation in the transformed VAR(1) is an AR(1) model. Therefore the results derived earlier in this subsection for the AR(1) model apply.

More generally, it should be noted that: 1) all of the variation in $\hat{A}_1^h$ is emanating from $\hat{\Lambda}_1^h$ (it was assumed E is known);

2)

$$var(vec[\hat{A}_1^h]) = var(vec[E\hat{\Lambda}_1^h E^{-1}]) = var([E'^{-1} \otimes E] vec[\hat{\Lambda}_1^h]) = [E'^{-1} \otimes E] var(vec[\hat{\Lambda}_1^h])[E'^{-1} \otimes E]'$$

Hence, the efficiency gains of impulse responses estimated via LP GLS impulse for a particular horizon depends on the relative efficiency of the eigenvalues, and how much an eigenvalue contributes to the variance of an impulse response. So if A_1 contains different eigenvalues, the eigenmatrices and the correlation among eigenvalues would determine how much the variance of an eigenvalue contributes to the variance of an impulse responses in the untransformed model and hence determine the relative efficiency of LP GLS impulse response to the VAR impulse responses. Essentially, the efficiency gains of the VAR come from the less persistent components. Depending on how many persistent eigenvalues there are and how much they contribute to the variance of the impulse responses, it is possible for LP GLS to be approximately as efficient as the VAR, when the true model is a VAR. Whether impulse responses of LP would be approximately as efficient would depend on the true data generating process, the persistence of the system, the dependence structure of the variables, and the horizon. In other words, it would be specific to the situation. It follows that the efficiency of the VAR relative to the LP has been overstated in the literature.

1.8 Monte Carlo Evidence

In this section I present Monte Carlo evidence of the finite sample properties of the bias adjusted LP GLS bootstrap. The properties of LP GLS estimator will analyzed along with the properties of the LP estimated via OLS with Newey-West standard errors (which will be referred to as LP NW), and the bias adjusted VAR bootstrap (Kilian, 1998). Bayesian LP will not be included in the Monte Carlo exercise because Bayesian methods do not carry the same interpretation as frequentist confidence intervals and their coverage cannot be assessed the same way (Rubin, 1984; Hoff, 2009).

The Monte Carlo simulations will deal with AR(1) models since it is easy to isolate the persistence and, as shown in the previous section, the results will be informative toward

VARs generally. The population model is

$$y_t = ay_{t-1} + \varepsilon_t,$$

where $a \in \{.99, .975, .95, .9, .75, .5\}$ and $\varepsilon_t \sim N(0, 1)$ and the sample size $T \in \{250\}$. The different values of a represent a range of eigenvalues encountered in macro. The sample size of 250 is representative of a quarterly data set dating back to 1960. Even though the most prominent macro variables such as GDP, inflation, and unemployment date back to at least 1948, many do not date back that far. The comprehensive McCracken and Ng (2016) data set goes back to 1959, so a sample size of 250 seems reasonable.

Simulations are conducted 1,000 times for each combination of a and T. For each simulation, I estimated the model for each desired horizon using all three methods and then check if the 95% confidence intervals contain the true impulse response. I then calculate the probability that the 95% confidence interval contains the true impulse response over the Monte Carlo simulations which gives me the coverage of the different methods. For each simulation draw, I also save the length of the 95% interval for the the different methods for each horizon. The lengths are then averaged over each Monte Carlo simulation for each method and horizon to get the respective average length of the 95% confidence intervals for each method and horizon. For the bias adjusted LP GLS and VAR bootstraps, I generate 1,000 bootstrap draws. I set the lag truncation parameter for the Newey-West standard errors to be $h - 1$, when estimating the horizon h LP. Note that for correctly specified VARs, this is the true lag truncation parameter. 15 horizons are analyzed, which would be representative of analyzing four years of impulse responses for quarterly data.

Figures A.1 and A.2 displays the coverage and average length respectively. In general the bias-adjusted LP bootstrap has good finite sample properties. It tends to have coverage at or near the nominal level, but coverage can decline somewhat at longer horizons, dropping

as low as 87%, if the autocorrelation coefficient is very persistent. This decline can be remedied to a degree by increasing the number of bootstraps. It is also the case that for the most persistent autocorrelation coefficients, coverage tends to be closer to 90% than 95%. The bias adjusted VAR bootstrap has coverage at or near the nominal level at all horizons. Consistent with the theoretical prediction in the previous section, the relative efficiency of the LP relative to the VAR depends on the persistence, with high persistence levels tending to have similar average lengths, which is consistent with asymptotic relative efficiency results in the previous section.

The LP NW estimator's performance can be much worse than the other two estimators. Coverage can drop drastically at higher horizons, below 60%, and if the data is persistent enough, coverage can be quite below the nominal level even at shorter horizons. The lack of coverage is due not only to the small sample bias but to Newey-West standard errors underestimating uncertainty. Note that the LP NW estimator tends to have shorter length than the bias adjusted LP GLS and VAR bootstraps. To "test" if this is due to Newey-West standard errors underestimating uncertainty, I estimate the "true" Monte Carlo variance of the different methods. That is, for each simulation I generated the AR(1) model and estimated the point estimates for each horizon using LP GLS, LP via OLS, and VAR via OLS. There is no bias correction here because the point is to show that even without bias correction, using GLS would lead to efficiency gains. To construct the VAR impulse responses, the OLS estimate was just raised to it's respective power. I saved the point estimates for each horizon for each simulation, and then I calculated the 95% quintiles (95% Monte Carlo confidence interval) for across the saved simulation estimates. This give me an approximation of the "true" 95% confidence intervals for these methods for the specific model and sample size. Figure A.3 displays the average length for the "true" 95% confidence intervals. The efficiency gains of using GLS are pretty clear and because the LP GLS and the VAR OLS do not use bias correction in this Monte Carlo, it follows that LP NW standard errors is underestimating uncertainty, and the lack of coverage is not solely due to the small sample bias. Newey-

West standard errors underestimating uncertainty is a common problem when the process is persistent (Müller, 2014). It is also important to note that Newey-West standard errors are underestimating uncertainty, even though the true lag truncation parameter is being used.

Even though the bias adjusted LP bootstrap displays good finite sample properties, the Monte Carlo analysis of the "true" confidence intervals also indicate there is some efficiency loss from using the block bootstrap. That is, the block bootstrap is not as efficient as it could be. The efficiency loss is probably due to the chosen block length, particularly the choice of consecutive blocks $L = T^{1/3}$. Additional Monte Carlos also show that changing the block length or implementing the stationary bootstrap can also improve the slight decline in coverage for the highly persistent processes, but there can also be a loss in efficiency. Since the block length involves a bias variance tradeoff with longer block lengths yielding less biased test statistics with larger variances and shorter block lengths yielding the opposite, a rule or a cross validation method such as Hall et al. (1995) needs to be developed.

1.9 Issues of Nonstationarity

It is also worth reiterating that the GLS procedure presented in section 2 and the consistency and asymptotic normality of the procedure assumes stationarity. Nonstationarity can be caused by unit roots or structural breaks. In the case of unit roots, inference from the frequentist perspective could differ depending on which variables have unit roots and what the parameters of interest are (Sims et al., 1990; Jordà, 2009). Consistency of the results can still hold if the errors have enough moments (Sims et al., 1990; Jordà, 2009), so the procedure can still eliminate autocorrelation, but asymptotic normality of the results could break down, so inference based on the frequentist procedures presented could be invalid if unit roots are present. If unit roots are an issue and the order of integration is known, the data could just be difference to stationarity. However the order of integration is probably not known. One could

test for unit roots, but frequentist tests unit roots lack power and can create considerable coverage distortions depending on the conclusion of the test (Pesavento and Rossi, 2006). In the case of Bayesian LP, Bayesian methods do not need to do anything special to take into account "explosive" nonstationarity behavior (e.g. unit roots) (Sims et al., 1990; Del Negro and Schorfheide, 2011), so estimation and inference involving Bayesian LP could proceed as usual.[22]

When nonstationarity is caused by structural breaks, both the frequentist and Bayesian methods presented will break down if they do not properly take into account change(s) in the parameters. Stationarity guarantees that the model has a linear time-invariant VMA representation. If the data are not stationary and structural breaks are the cause, then the procedure may not eliminate autocorrelation. To understand why it matters if structural breaks are present, note that if the data are not stationary, it is possible for the estimated horizon 0 LP residuals to be uncorrelated since the VAR can still produce reasonable one-step ahead forecasts when the model is misspecified (Jordà, 2005). A "Wold representation" exists for nonstationary data, but the impulse responses for this VMA representation are allowed to be time dependent (Granger and Newbold, 1977; Priestley, 1988).[23] Assuming there is no deterministic component, any time series process can be written as

$$y_t = \varepsilon_t + \sum_{i=1}^{\infty} \Theta_{i,t} \varepsilon_{t-i},$$

where $\Theta_{i,t}$ is now indexed by the horizon and time period and $var(\varepsilon_t) = \Sigma_{\varepsilon,t}$. Using recursive substitution, the time dependent Wold representation can be written as a time dependent VAR or a time dependent LP.[24] It can be shown that a time dependent version of Theorem

[22]There is a lively debate about how to construct priors for Vector Error Correction models (Del Negro and Schorfheide, 2011).

[23]Nonstationarity in economics typically refers to explosive behavior (e.g. unit roots), but nonstationarity is more general and refers to a distribution that does not have a constant mean and/or variance over time (e.g. threshold models or models with stochastic volatility). Depending on the true model, differencing may not make the data stationarity (Leybourne et al., 1996; Priestley, 1988).

[24]The lag lengths can be infinite. Obviously in practice, a finite lag length would be chosen.

1 exists. The horizon h time dependent LP is

$$y_{t+h} = B_{1,t}^{(h+1)} y_{t-1} + B_{2,t}^{(h+1)} y_{t-2} + \ldots + e_{t+h}^{(h)}, \tag{1.10}$$

where

$$e_{t+h}^{(h)} = \Theta_{h,t}\varepsilon_t + \ldots + \Theta_{1,t}\varepsilon_{t+h-1} + \varepsilon_{t+h}$$

$$B_{1,t}^{(h)} = \Theta_{h,t}.$$

If impulse responses are time dependent at higher horizons, but a time invariant version of LP GLS is applied, autocorrelation may not be eliminated at these horizons because the time-invariant LP are misspecified. In other words, if the data are nonstationary and the nonstationarity is caused by structural breaks, the time invariant version of LP GLS may not eliminate autocorrelation in the residuals since the estimates of the impulse responses may not be consistent. In this sense, LP GLS is a type of general misspecification test, because if one had estimated LP using OLS and Newey-West standard errors, the autocorrelation in the residuals would not hint toward potential misspecification since the residuals are inherently autocorrelated.

As noted in Granger and Newbold (1977), macro data encountered in practice are unlikely to be stationary, implying that the Wold representation may be time dependent. If the impulse responses of the Wold representation are time dependent, since time-varying parameter models can approximate any form of non-linearity (Granger, 2008), a time varying version of LP GLS may be applied. The time-varying parameter version of the above GLS procedure presented in section 1.3 will be able to eliminate autocorrelation as long as the parameter changes are not so violent that a time-varying parameter model cannot track them. All else equal, the more adaptive the time-varying parameter model, the better the time-varying

parameter model will be able to track changes and the better the approximation.[25] Time-varying parameter LP are presented in the next section. If the nature of the time dependence is known, that is, the researcher knows when the structural breaks occur or the nature of the time variation (i.e. regime switching models for expansions and recessions), then that specific time dependent model can be applied to the LP GLS procedure.

1.10 Time-Varying Parameter LP

As noted in the introduction, a researcher may be interested in allowing for time-varying parameters. Stock and Watson (1996) and Ang and Bekaert (2002) show many macroeconomic and financial time series exhibit parameter instability. It is also commonplace for regressions with macroeconomic time series to display heteroskedasticity of unknown form (Stock and Watson, 2007a), and in order to do valid inference, the heteroskedasticity must be taken into account. Parameter instability can occur for many reasons such as policy changes, technological evolution, changing economic conditions, etc. If parameter instability is not appropriately taken into account, it can lead to invalid inference, poor out of sample forecasting, and incorrect policy evaluation. Moreover, time-varying parameter models can approximate any non-linear model (non-linear in the variables and/or the parameters), which makes them more robust to model misspecification (Granger, 2008).

As mentioned in the previous section, for any time series process, there exists a time dependent Wold representation

$$y_t = \varepsilon_t + \sum_{i=1}^{\infty} \Theta_{i,t} \varepsilon_{t-i},$$

where $\Theta_{i,t}$ is now indexed by the horizon and time period and $var(\varepsilon_t) = \Sigma_{\varepsilon,t}$. Using recursive

[25] Baumeister and Peersman (2012) show via Monte Carlo simulations that time-varying parameter models are able to capture discrete breaks in a satisfactory manner should they occur.

substitution, the time dependent Wold representation can be written as a time dependent VAR or a time dependent LP. It can be shown that a time dependent version of Theorem 1 exists. The horizon h time dependent LP is

$$y_{t+h} = B_{1,t}^{(h+1)} y_{t-1} + B_{2,t}^{(h+1)} y_{t-2} + \ldots + B_{k,t}^{(h+1)} y_{t-k} + e_{t+h}^{(h)}, \tag{1.11}$$

where

$$e_{t+h}^{(h)} = \Theta_{h,t}\varepsilon_t + \ldots + \Theta_{1,t}\varepsilon_{t+h-1} + \varepsilon_{t+h}$$

$$B_{1,t}^{(h)} = \Theta_{h,t}.$$

Just like the time invariant case, k can be infinite in population but will be truncated to a finite value in finite samples. Similarly to the time-invariant transformation, one can do a GLS transformation $\widetilde{y}_{t+h}^{(h)} = y_{t+h} - \hat{B}_{1,t}^{(h)}\hat{\varepsilon}_t - \ldots - \hat{B}_{1,t}^{(1)}\hat{\varepsilon}_{t+h-1}$. Then one can estimate horizon h via the following equation

$$\widetilde{y}_{t+h}^{(h)} = B_{1,t}^{(h+1)} y_{t-1} + B_{2,t}^{(h+1)} y_{t-2} + \ldots + B_{k,t}^{(h+1)} y_{t-k} + \tilde{u}_{t+h}^{(h)}. \tag{1.12}$$

Estimation is carried out in the same way as in the time-invariant case, except the models are being estimated with time-varying parameters.

Just like a static LP model can be more robust to model misspecification than a static VAR, a time-varying parameter LP model can be more robust to model misspecification than a time-varying parameter VAR. If the true model is time varying, then the misspecification of the VAR can extend to the time variation as well. Due to the iterative nature of the VAR, misspecification in time variation would be compounded in the construction of the impulse responses alongside other misspecifications in the VAR. Time-varying parameter LP, however, allow for the amount and nature of time variation to change across horizons. Since

time-varying parameter models can also approximate any non-linear model, time-varying parameter LP can do a to better job capture the time variation in the impulse responses at each horizon.

There are several ways to estimate time-varying parameter models. Bayesian methods are the primary methods used to estimate time-varying parameter models, and because autocorrelation is explicitly corrected for in Bayesian LP, it is straightforward to apply time-varying parameters to Bayesian LP. For the rest of this section, I will describe a computationally convenient way to estimate time-varying parameter models. This procedure is based off of Prado and West (2010). Let

$$y_t = X_t'\beta_t + v_t,$$

$$\beta_t = \beta_{t-1} + w_t,$$

where y_t is a $r \times 1$ vector β_t is the $p \times 1$ state vector at time t, X_t is a $p \times r$ vector of regressors at time t, v_t is a $r \times 1$ vector observation noise with $v_t \sim N(0, \Sigma_t)$, w_t is the state evolution noise with $w_t \sim N(0, \Sigma_t \otimes W_t)$, and v_s and w_t are independent and mutually independent $\forall s, t$. Notice that the variance of v_t is allowed to be time-varying. Stochastic volatility (time-varying variance) is modeled as a beta-Bartlett Wishart random walk. Define D_{t-1} is the amount of information known at time $t-1$. The beta-Bartlett Wishart random walk is defined using the following $t-1$ to time t update

$$p(\Sigma_{t-1}|D_{t-1}) \sim IW(n_{t-1}, \Psi_{t-1})$$

and

$$p(\Sigma_t|D_{t-1}) \sim IW(\theta n_{t-1}, b_t \Psi_{t-1}),$$

where θ is a discount factor for stochastic volatility and $b_t = (\theta n_{t-1} + k - 1)/(n_{t-1} + k - 1)$.[26] The models are estimated using discount factors and the Forward Filter Backward Sampler (FFBS) algorithm, and details about the estimation procedure can be found in the Appendix.[27] Because discount factors and conjugate priors are used, MCMC is not needed. This is crucial for three reasons. First, if the number or parameters is even moderately large, time-varying parameter models such as Cogley and Sargent (2005); Primiceri (2005) become computationally demanding to estimate if not infeasible (Koop and Korobilis, 2013). Second, LP are estimated horizon by horizon in a sequential fashion which can make procedures such as Cogley and Sargent (2005); Primiceri (2005) impractical. Third, in order to do model comparison or hypothesis testing, it is often necessary to calculate the marginal likelihood, which is no trivial task for models estimated using MCMC. In recent years discount factors have been used in the as a solution to when the procedures of Cogley and Sargent (2005); Primiceri (2005) are burdensome (Koop and Korobilis, 2013; Koop et al., 2018). This is not to suggest that time-varying parameter procedures such as Cogley and Sargent (2005); Primiceri (2005) or other cannot be used, just that depending on the goal of the analysis and the computational power available to the researcher, these procedures may not be practical.[28]

Discount factors (also known as forgetting factors) are a natural framework for allowing and controlling for time variation in regression coefficients and the variance and are a core part of the Bayesian forecasting literature (West and Harrison, 1997; Prado and West, 2010). Discount factors lie in the interval $(0, 1]$. If a discount factor, say $\theta = .99$ is used, then from period $t \to t+1$, $\frac{1}{\theta} - 1 \approx 1\%$ of information known at time t is discounted or forgotten in the Kalman filtration process.[29] And if $\theta = .99$, observations from 20 periods ago receive

[26]The model uses different discount factors for the regression coefficients and stochastic volatility.

[27]See West and Harrison (1997); Prado and West (2010) for derivations and more details about time-varying parameter methods using discount factors.

[28]If time-varying parameter procedures such as Cogley and Sargent (2005); Primiceri (2005) are used, it is recommended that the MCMC be implemented using the more computationally efficient precision sampler in Chan and Jeliazkov (2009).

[29]A discount factor of .99 has properties similar to what Cogley and Sargent (2005) call their "business as usual" prior, and it can be shown that the choice of prior shrinkage coefficient in Cogley and Sargent (2005) allows for variation in the regression coefficients roughly similar to that allowed for by a regression

approximately 80% as much weight as this period's observation. The loss of information over time allows more recent data to have a larger impact on the parameter value and is the crux for controlling for time variation in the parameters. The discount factors are estimated using Griddy Gibbs. Including the the discount factor as a parameter to be estimated takes into account uncertainty in the hyperparameters and is a natural way to safeguard against overfitting (Giannone et al., 2015).

Due to the number of parameters being estimated, the priors for time-varying parameter models are quite important (Koop and Korobilis, 2009), otherwise parameter estimates may be imprecise if the sample size is not large. Like Cogley and Sargent (2005); Primiceri (2005), a training sample prior can be used. The prior is the same as the one presented earlier in section 1.5.[30]

1.11 Concluding Remarks

I show that LP can be estimated with GLS. Estimating LP with GLS has three major implications. First, LP GLS can be substantially more efficient and less biased than estimation by OLS with Newey-West standard errors. Moreover, if the data are persistent and the true model is a VAR, it can be shown that impulse responses from LP can be approximately as efficient as impulse responses from VARs. Whether or not the LP is approximately as efficient depends on the persistence of the system, the horizon, and the dependence structure of the system. All else equal, the more persistent the system, the more likely LP impulse responses will be approximately as efficient for horizons typically relevant in practice. Given that most macro data are nonstationary or nearly nonstationary, even if the true model is a VAR, the efficiency of the VAR relative to the LP has been overstated in the literature.

coefficient discount factor of .99 (Koop and Korobilis, 2013).

[30]It should be noted that non-informative priors (such as reference priors) cannot be used in Bayesian model comparison due to Bartlett's paradox. If a training sample is not available, other priors can be used. See Koop and Korobilis (2009); Koop (2017) for a review.

Second, because autocorrelation process can be modeled explicitly, it is possible to give a fully Bayesian treatment of LP. That is, LP can be estimated using fully Bayesian or frequentist methods. Bayesian LP have many advantages over frequentist LP and/or Bayesian VARs such as allowing the researcher to incorporate prior information for impulse responses at each horizon. Prior information can be used to shrink impulse responses at any horizon to prevent overfitting. Economic theory can be incorporated into the prior to inform the shape of the impulse responses (e.g. the impulse response is monotonic or hump shaped) and to discipline the long-run behavior. Bayesian methods do not need to do anything special to take into account nonstationarity.

Third, since autocorrelation process can be modeled explicitly, it is now possible to estimate time-varying parameter LP. Bayesian LP can easily be adapted to handle time-varying parameter models, but one does not have to use Bayesian methods. Time-varying parameter LP can take into account structural instability in the regression coefficients and/or the covariance matrix, and since time-varying parameter models can approximate any form of non-linearity, makes them more robust to model misspecification (Granger, 2008).

The results in this paper have many potential extensions for both frequentist and Bayesian analysis. It would be useful for frequentist to have a data dependent rule or cross validation method for the optimal block length when using block bootstrapping for LP. It may be useful to extend some of the big data, sparsity, and variable selection methods used for VARs to LP.[31] It may also be useful to extend LP GLS to a non-linear (in the variables) or non-parametric setting. Even though time-varying parameter models can approximate any non-linear model (non-linear in the variables and/or the parameters), the approximation is for the conditional mean, so if the true model is non-linear in the variables, estimation of the linear (in the variables) time-invariant or time-varying parameter LP GLS would lead to inconsistent estimates of the true impulse responses. One potential solution would be to

[31] See Koop and Korobilis (2009); Koop (2017) for a review.

extend polynomial LP, which are motivated by non-linear version of the Wold representation (see Jordà (2005) section 3 for more details). If one is does not want to make assumptions about the functional form or the model, the second potential solution would be to extend nonparametric LP. Lastly, since LP are direct multistep forecasts, the results in this paper have the potential to improve the forecast accuracy of direct multistep forecasts.

Chapter 2

Identifying Government Spending Shocks, Fiscal Foresight, and Time-Varying Parameters

The nature of the financial crisis of 2007-2009 limited the Federal Reserve's ability to induce demand through monetary policy. Many economists and policy makers advocated fiscal stimulus to prevent a depression, and in 2009, the American Recovery and Reinvestment Act (ARRA) was passed and enacted into law. The subsequent recovery has been considered to be tepid by economists, lawmakers, and the general public. The nature of the crisis and the tepid recovery has renewed a nearly century's old debate about the impact of short-run government fiscal policy on the economy. Many have argued that the stimulus prevented a depression, while many others have argued that it was an inefficient use of government funds that just crowded out spending in the private sector. At the heart of the debate is the question: What is the size of the government expenditure multiplier?[1] As noted in Ramey

[1]Government spending and expenditures will be used interchangeably throughout the paper. Fiscal, spending, and expenditure multipliers will also be used interchangeably.

(2011a, 2019), there are a wide range of estimates for multipliers in both empirical studies and dynamic simulations.

Since the Great Recession, there has been a "renaissance" in fiscal research and there have been several innovations made in the literature. One of the most important innovations has been the recognition of fiscal foresight. That is, the recognition that a government may announce fiscal changes in advance, and if these announcements are not somehow taken into account in the model, multiplier estimates can be severely biased. This was formally shown in Leeper et al. (2013), who argue that fiscal foresight is an omitted variable bias problem, and the solution to the fiscal foresight is to use an expectations-augmented model. Since then, essentially all fiscal multiplier papers try to address fiscal foresight through narrative methods or by augmenting there models with forecasts of future government spending (Ramey, 2019).[2] I will show that fiscal foresight is not an omitted variable (invertibility) problem. The solution to fiscal foresight depends on the proposed fiscal rule. Once a more natural rule is used, I show both theoretically and empirically that the solution to the fiscal foresight problem is not to use expectation-augmented models, but to use time-varying parameter models or models that allow for structural breaks.

Another important innovation made in fiscal research in the last decade is calculating multipliers in a dynamic environment. The "natural" way people think of multipliers is in terms of dollars: A dollar-based multiplier of 1 means that if government increases spending by a dollar, GDP would increase by a dollar. Since the seminal study of Blanchard and Perotti (2002), it has been noted in several papers that since the GDP-to-spending ratio varies over time, converting estimated elasticities into dollars after estimation can be difficult. A solution that was proposed and is widely adopted is to convert GDP and spending to the same units before estimation, so ex post conversion is not needed. I will show that converting GDP and spending to the same units before estimation may have biased multiplier estimates

[2]Some studies treat the expectation as the actual shock while others use it as an instrument.

of previous studies. Moreover, I will explain why the multiplier should not be expressed in terms of dollars generally, and why the dollar based multiplier would not actually tell us the impact that government spending has on the economy.

This paper makes two major contributions. First, I show theoretically and empirically that fiscal foresight is not an omitted variable problem, and I explain why time-varying parameters models are a solution to the problem. Second, I explain why not estimating fiscal multipliers using time-varying parameter models, and bit expressing fiscal multipliers in dollars can severely bias multiplier estimates and lead to incorrect inference.

2.1 The Role of Fiscal Foresight

To illustrate why fiscal foresight is not an omitted variable problem, I will first show that changing the fiscal rule in Leeper et al.'s (2013) canonical tax foresight example leads to a different solution for fiscal foresight. I will then show how a similar argument applies for spending foresight. Lastly, I will show how the empirical evidence of fiscal foresight in Ramey (2011b); Auerbach and Gorodnichenko (2012) do not hold.

2.2 Tax Foresight Example and Critique

To explain why tax foresight is not an omitted variable problem, I will explain how tax foresight works using the same model as Leeper et al. (2013). I will then show why making the rule more realistic changes the solution to tax foresight. Consider a standard growth model with a representative household that maximizes expected utility

$$max\, E_0 \sum_{t=0}^{\infty} \beta^t log(C_t),$$

$$s.t. C_t + K_t - T_t = (1-\tau_t)A_t K_{t-1}^{\alpha},$$

where C_t, K_t, Y_t, T_t, and τ_t, are the time t consumption, capital, output, lump-sum transfers, and the income tax rate, respectively, and A_t is an exogenous technology shock. As usual, $0 < \alpha < 1$ and $0 < \beta < 1$. The government sets the tax rate according to a time-invariant rule and adjusts lump sum transfers to satisfy the constraint $T_t = \tau_t Y_t$. Government spending is identically zero. There is complete depreciation of capital. Labor is supplied inelastically.

The equilibrium conditions are well known and given by

$$\frac{1}{C_t} = \alpha\beta E_t \left[(1-\tau_{t+1}) \frac{1}{C_{t+1}} \frac{Y_{t+1}}{K_t} \right],$$

$$C_t + K_t = Y_t = A_t K_{t-1}^{\alpha}.$$

Let A and τ denote the steady-state values of technology and the tax rate. The steady-state capital stock is $K = [\alpha\beta(1-\tau)A]^{1/(1-\alpha)}$. Let lower-case letters denote percentage deviations from steady-state values, $k_t = log(K_t) - log(K)$, $a_t = log(A_t) - log(A)$, and $\hat{\tau}_t = log(\tau_t) - log(\tau)$. Log-linearizing and combining the two equilibrium conditions yields a second-order difference equation in capital,

$$E_t k_{t+1} - (\theta^{-1} + \alpha)k_t + \alpha\theta^{-1}k_{t-1} = E_t[a_{t+1} - \theta^{-1}a_t] + \{\theta^{-1}(1-\theta)(\frac{\tau}{1-\tau})\}E_t\hat{\tau}_{t+1},$$

where $\theta = \alpha\beta(1-\tau)$. Assuming an i.i.d. technology shock, the solution for the above equation is

$$k_t = \alpha k_{t-1} + a_t - (1-\theta)(\frac{\tau}{1-\tau}) \sum_{i=0}^{\infty} \theta^i E_t \hat{\tau}_{t+i+1}.$$

Equilibrium investment depends negatively on the expected discounted present value of future tax rates, which is a well-known result. More distant tax rates receive heavier discounting than less distant rates.

To model foresight, one must specify how news about taxes signals future tax rates. Leeper et al. (2013) assume the tax information flows take a particularly simple form: agents at time $t-q$ receive a signal that tells them exactly what tax rate they will face in period t. The tax rule is

$$\tau_t = \tau exp(\varepsilon_{\tau,t|t-q}),$$

or in log-linearized form

$$\hat{\tau}_t = \varepsilon_{\tau,t|t-q}.$$

$\varepsilon_{\tau,t|t-q}$ is a tax shock that will occur at time t but the agent has known about since time $t-q$. It is assumed that the technology and tax shocks– $\varepsilon_{A,t}$ and $\varepsilon_{\tau,t|t-q}$–are i.i.d. and the representative agent's information set at time t consists of variables known at time t , including the shocks, $\{\varepsilon_{A,t}, \varepsilon_{\tau,t|t-q}, \ldots, \varepsilon_{\tau,t+q|t}\}$. Given the tax news process, this implies that at time t, the agent has (perfect) knowledge of $\{\hat{\tau}_{t+q}, \hat{\tau}_{t+q-1}, \ldots\}$.

Using the information flows to solve for expected tax rates for various degrees of fiscal foresight yields the following equilibrium dynamics:

$q = 0$ implies:

$k_t = \alpha k_{t-1} + \varepsilon_{A,t};$

$q = 1$ implies:

$k_t = \alpha k_{t-1} + \varepsilon_{A,t} - \kappa \varepsilon_{\tau,t+1|t};$

$q = 2$ implies:

$k_t = \alpha k_{t-1} + \varepsilon_{A,t} - \kappa \{\varepsilon_{\tau,t+1|t-1} + \theta \varepsilon_{\tau,t+2|t}\};$

$q = 3$ implies:

$$k_t = \alpha k_{t-1} + \varepsilon_{A,t} - \kappa\{\varepsilon_{\tau,t+1|t-2} + \theta\varepsilon_{\tau,t+2|t-1} + \theta^2\varepsilon_{\tau,t+3|t}\};$$

where $\kappa = (1-\theta)(\tau/(1-\tau))$.

If there is no foresight ($q = 0$) we get the usual result that i.i.d. shocks to tax rates have no effect on capital accumulation. When there is some degree of tax foresight ($q > 0$), rational agents will adjust capital contemporaneously to yield the result that even serially uncorrelated tax hikes reduce capital accumulation. Fiscal foresight manifests in the additional moving average terms present in the equilibrium representation, with the number of moving average terms increasing in the foresight horizon.

Fiscal foresight is a problem for the econometrician because if one wants to recover the technology shock $\varepsilon_{A,t}$ by regressing k_t on lags of itself, one will be unable to identify $\varepsilon_{A,t}$ because the residual from the autoregressive model is not unexpected and will be a mixture of news that is already known to the agent and $\varepsilon_{A,t}$. To solve this problem, one would need to augment the autoregressive model with news of future tax shocks. That is, one would need to use an Expectations Augmented VAR.

I argue that the information flows would not come in the form of "shocks". In Leeper et al. (2008), multiple examples are used to motivate fiscal foresight (e.g. tax reform of 1986, Bush tax cuts, The Economic Recovery Tax Act of 1981). All of these examples involve anticipation of changes in the tax code or the announcement of changes in the tax code. Since changes in the tax code are changes in the "rules", I argue that treating news as a shock does not properly capture foresight. If there is a tax rule, and the tax code determines the parameters in the rule, a change in the tax code will cause the parameters in the tax rule to change. I posit a more appropriate tax rule is

$$\tau_t = \tau + \varepsilon_{\tau,t},$$

where again τ is the steady state (and mean) tax rate and $\varepsilon_{\tau,t}$ is an unexpected tax shock. If it is announced today that the tax code is going to change permanently q periods from now, then the intercept in the rule should increase (or decrease) permanently to τ' q periods from now, and the rule becomes

$$\tau_{t+q} = \tau' + \varepsilon_{\tau,t+q}.$$

This is a standard way to treat an announcement of a policy change (Ljungqvist and Sargent, 2012). This rule has the advantage that agents don't have to know the future tax rate with 100% certainty (since it still allows for unexpected random shocks), but the change in the future tax rate is still allowed to affect expectations. Since agents recognize that there is a new steady state, note that the deviation from the steady state for $j < q$ is

$$\hat{\tau}_{t+j} = log(\tau_{t+j}) - log(\tau') = log(\tau) - log(\tau') + \varepsilon_{\tau,t+j},$$

and the time t expectation of $\hat{\tau}_{t+j}$ is

$$E_t\hat{\tau}_{t+q} = log(\tau) - log(\tau').$$

Investment at time $t+j$ is still affected by the announcement, but once at time $t+q$, the announcement no longer affects capital. Fiscal foresight is still a problem, but this problem cannot be solved by estimating an expectations augmented VAR. To see this, recall that the equilibrium solution to capital is

$$k_t = \alpha k_{t-1} + a_t - (1-\theta)(\frac{\tau'}{1-\tau'})\sum_{i=0}^{\infty}\theta^i E_t\hat{\tau}_{t+i+1}.$$

Since

$$E_t\hat{\tau}_{t+j} = log(\tau) - log(\tau'),$$

for $j < q$, augmenting an autoregressive model of capital with expectations of future tax rates does not fix the problem because the announcement of the change from τ to τ' causes the intercept in k_t to be time-varying, as opposed to causing k_t to have a non-invertible representation.

If the announcement is for a temporary change, say q periods from now, the rule changes from

$$\tau_t = \tau + \varepsilon_{\tau,t},$$

to

$$\tau_t = \tau' + \varepsilon_{\tau,t},$$

for h periods, the above argument still holds, but it should be noted that the steady state will not shift to τ' because agents know the change is temporary. Therefore for $j = q, q + 1, \ldots, q + h$

$$E_t \hat{\tau}_{t+j} = log(\tau') - log(\tau),$$

and

$$k_t = \alpha k_{t-1} + a_t - (1 - \theta)(\frac{\tau}{1 - \tau}) \sum_{i=0}^{\infty} \theta^i E_t \hat{\tau}_{t+i+1}.$$

The rule can even be augmented to handle both announcements of changes in the tax code and news that the tax code may change. Say there's news today (e.g. Trump being elected) that taxes may change in the future. The agent is not sure of exactly how the election will

affect taxes but has a prior for the effect. The tax rule can be written as

$$\tau_{t+q} = \tau + m_{t+q} + \varepsilon_{\tau,t+q}$$

where $E[m_{t+q}] = \mu_q$ and the $Variance[m_{t+q}] = \sigma_q^2$. m_{t+q} is a random variable from the prior distribution of the effect the agent thinks Trump's election will have on taxes in period $t+q$ (the distribution of m_{t+q} may have to be truncated to ensure that the tax rate is above zero but below 1). If an agent suspects that the election of Trump will lower taxes, μ_{t+q} will be negative.

$$\hat{\tau}_{t+q} = log(\tau_{t+q}) - log(\tau) = log(\tau) + m_{t+q} + \varepsilon_{\tau,t+q} - \tau.$$

And the time t expectation of τ_{t+q} is

$$E_t\hat{\tau}_{t+q} = log(\tau) + \mu_{t+q}.$$

Like for the announcement of a change, news of a potential change causes the intercept in the capital equation to be time-varying, so an expectation-augmented model would not fix the problem. This rule has the advantage of the agents only need to have a prior for how they think news might affect the tax rule; they do not have to know with certainty, which is likely to be the case in practice. Furthermore, if there is an announcement of actual changes, the agent would just adjust τ.

2.3 Spending Foresight Example

The results for taxes foresight can be extending to spending as well. This example is used in Perotti (2011). Again, consider a standard growth model with a representative household

that maximizes expected utility

$$max\, E_0 \sum_{t=0}^{\infty} \beta^t log(C_t),$$

$$s.t.\, C_t + K_t - T_t = (1-\tau_t) A_t K_{t-1}^{\alpha},$$

$$0 < \alpha < 1,\, 0 < \beta < 1,\, T_t = G_t,$$

where C_t, K_t, Y_t, G_t, and T_t are the time t consumption, capital, output, government spending, and lump sum transfers, respectively, and A_t is an exogenous technology shock. We assume complete depreciation of capital. Labor is supplied inelastically.

After log-linearizing the Euler equation and the resource constraint, the solution to capital is

$$k_t = \lambda_1 k_{t-1} + \frac{1}{\lambda_2 \alpha \beta} a_t - \frac{G}{\lambda_2 \alpha \beta} \sum_{i=0}^{\infty} \frac{1}{\lambda_2^i} E_t(\hat{g}_{t+1+i} - \hat{g}_{t+1}),$$

where $\lambda_1 < 1$ and $\lambda_2 > 1$ are the roots of the characteristic equation.[3] The standard rule used to illustrate fiscal foresight for government spending process is

$$G_t = Gexp(\varepsilon_{G,t|t-q} + \varepsilon_{G,t}),$$

or in log deviation form

$$\hat{g}_t = \varepsilon_{G,t|t-q} + \varepsilon_{G,t}.$$

If agents have 1 period of foresight ($q = 1$) then capital will adjust contemporaneously to

[3]See Perotti (2011) for explicit definitions of λ_1 and λ_2.

news

$$k_t = \lambda_1 k_{t-1} + \delta\varepsilon_{A,t} + \pi_0\varepsilon_{G,t|t-1} + \pi_1\varepsilon_{G,t+1|t} + \pi_0\varepsilon_{G,t}.$$

[4] For example, if $q = 1$, and an econometrician were to estimate a VAR of spending and capital then

$$\begin{bmatrix} \hat{g}_t \\ k_t \end{bmatrix} = intercept + \sum_{j=1}^{p} \Phi_j \begin{bmatrix} \hat{g}_{t-j} \\ k_{t-j} \end{bmatrix} + \varepsilon_t,$$

and

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{G,t|t-1} + \varepsilon_{G,t} \\ \delta\varepsilon_{A,t} + \pi_0\varepsilon_{G,t|t-1} + \pi_1\varepsilon_{G,t+1|t} + \pi_0\varepsilon_{G,t} \end{bmatrix}.$$

A VAR of spending and capital will yield residuals that are predictable with information outside of the VAR's information set, so structural shocks cannot be recovered from a reduced form VAR of spending and capital using SVAR methods. The solution would be to use an Expectations Augmented VAR (EVAR) and include expectations of future government spending as an additional variable.

Just like the tax example, treating foresight as shocks does not align with the examples used to motivate fiscal foresight. If legislation determines the parameters in the fiscal rule, then a change in the legislation would change the parameters in the fiscal rule. I posit a more appropriate spending rule is

$$log(G_t) = log(G) + \varepsilon_{G,t},$$

where G is the steady state spending level and $\varepsilon_{G,t}$ is an unexpected spending shock. If it's

[4] See Perotti (2011) for explicit definitions of δ, π_0, and π_1.

announced today that there is going to be a permanent change in government spending q periods from now, I argue that the new rule would be

$$log(G_{t+q}) = log(G') + \varepsilon_{G,t+q},$$

where G' is the new steady state for government spending. Again, this is a standard way to treat an announcement of a policy change (Ljungqvist and Sargent, 2012).

Since agents recognize that there is a new steady state, note that the deviation from the steady state for $j < q$ is

$$\hat{g}_{t+j} = log(G_{t+j}) - log(G') = log(G) - log(G') + \varepsilon_{G,t+j},$$

and the time t expectation of $\hat{g}_{t+j}$ is

$$E_t\hat{g}_{t+j} = log(G) - log(G').$$

Investment at time $t + j$ is still affected by the announcement, but once at time $t + q$, the announcement no longer affects capital. Fiscal foresight is still a problem, but this problem cannot be solved by estimating an expectations augmented VAR. The new solution to capital is

$$k_t = \lambda_1 k_{t-1} + \frac{1}{\lambda_2\alpha\beta}a_t - \frac{G'}{\lambda_2\alpha\beta}\sum_{i=0}^{\infty}\frac{1}{\lambda_2^i}E_t(\hat{g}_{t+1+i} - \hat{g}_{t+1}).$$

Since the time t expectation of $\hat{g}_{t+j}$ for $j = 1, 2, \ldots$ is either $log(G) - log(G')$ or 0, the intercept in the equilibrium equation is time-varying and depends on the specific t.

If one were to estimate a VAR with a time-varying intercept (or an intercept that can handle

structural breaks) for spending and capital, then

$$\begin{bmatrix} \hat{g}_t \\ k_t \end{bmatrix} = intercept_t + \sum_{j=1}^{p} \Phi_j \begin{bmatrix} \hat{g_{t-j}} \\ k_{t-j} \end{bmatrix} + \varepsilon_t,$$

and

$$\varepsilon_t = \begin{bmatrix} \varepsilon_{G,t} \\ \delta\varepsilon_{A,t} + \pi_0\varepsilon_{G,t} \end{bmatrix}.$$

A VAR of spending and capital will yield residuals that are unexpected so structural shocks can be recovered from a reduced form VAR using SVAR methods (assuming the econometricians restrictions are valid).

In the case of a temporary change in government spending, the rule would be

$$log(G_t) = log(G) + \varepsilon_{G,t},$$

where G is the steady state spending level and $\varepsilon_{G,t}$ is an unexpected spending shock. If it's announced today that there is going to be a temporary change in government spending q periods from now and the change will last for h periods, then for $j = q, q+1, \ldots, q+h$

$$log(G_{t+j}) = log(G') + \varepsilon_{G,t+j},$$

where G' is the temporary "level" for government spending. Since the change is not permanent, the steady state does not change. For $j = q, q+1, \ldots, q+h$

$$\hat{g}_{t+j} = log(G_{t+j}) - log(G) = log(G') - log(G) + \varepsilon_{G,t+j},$$

and the time t expectation of $\hat{g}_{t+j}$ is

$$E_t\hat{g}_{t+j} = log(G') - log(G).$$

Again, fiscal foresight is still a problem, but the problem can be solved using a time-varying parameter VAR.

This rule can also handle news that the future government spending may change. Say there's news today (e.g. Trump being elected) that government spending may change in the future. The agent is not sure of exactly how the election will affect spending but has a prior for the effect. The spending rule can be written as

$$log(G_{t+q}) = log(G) + m_{t+q} + \varepsilon_{G,t+q},$$

where $E_t[m_{t+q}] = \mu_{t+q}$ and the $Variance_t[m_{t+q}] = \sigma^2_{t+q}$. $Pr(m_{t+q})$ is a prior distribution of the effect that an agent thinks Trumps's election will have on spending in period $t+q$. If an agent suspects that the election of Trump will reduce future spending, μ_{t+q} will be negative.

$$\hat{g}_{t+q} = log(G_{t+q}) - log(G) = m_{t+q} + \varepsilon_{G,t+q}.$$

And the time t expectation of $\hat{g}_{t+q}$ is $E_t\hat{g}_{t+q} = \mu_{t+q}$. Like for the announcement of a change, news of a potential change causes the intercept in the capital equation to be time-varying, so an expectation-augmented model does not fix the problem. This rule has the advantage that agents only need a prior for how they think the news might affect the spending rule; they don't have to know with certainty.

2.4 Empirical Evidence

The theoretical results of Leeper et al. (2013) have been used to motivate arguments as to why the Blanchard and Perotti (2002) shock is predictable. The reasoning for this goes as follows: the Blanchard-Perotti shock is the error term in the government spending equation in a VAR where government spending is ordered first. Since a VAR is only conditional on past information of the variables in a VAR, a VAR cannot take into account news or announcements unless a news/expectations variable is explicitly embedded into the VAR. The predictability of the Blanchard-Perotti shock could lead to downward bias in the multiplier estimates because the full reaction to an unanticipated shock is not being captured.[5] The two most prominent studies that provide evidence that the Blanchard-Perotti shock is predictable are Ramey (2011b) and Auerbach and Gorodnichenko (2012). Ramey (2011b) shows the Blanchard-Perotti shock is predictable using a defense news variable that captures anticipations in future military spending. Alternatively, Auerbach and Gorodnichenko (2012) show the Blanchard-Perotti shock is predictable using professional forecast of government spending.[6]

I will show that the aforementioned empirical evidence in Ramey (2011b); Auerbach and Gorodnichenko (2012) is overstated for two different reasons. The first reason is that their models do not allow for time-varying parameters. The theoretical justification for using using time-varying parameters was given in the previous subsections. The second, more subtle reason, is that their tests may be subject to the generated regressor problem (Pagan, 1984). Ramey (2011b) first generates the Blanchard-Perotti shock in a first step, and then does Granger causality tests in a second step, but Ramey's (2011b) Granger causality tests do not take into account that the the Blanchard-Perotti shocks are estimated. Auerbach and

[5] Ramey (2009) showed in a stylized neoclassical growth model that missing the timing of news mutes the size of the multiplier because the anticipatory rise in GDP is not captured.

[6] Ramey (2011b) also uses professional forecast of government spending, but Auerbach and Gorodnichenko (2012) forecast of government variable covers a longer time span, which includes the time span covered in Ramey (2011b).

Gorodnichenko (2012) estimate Blanchard-Perotti shocks for government spending and for forecasts of government spending, and show that the two sets of Blanchard-Perotti shocks are correlated, which would imply that the standard Blanchard-Perotti shock is predictable and can be further purged of information available at the time of the forecast. Auerbach and Gorodnichenko's (2012) test does not take into account that the Blanchard-Perotti shocks are estimated. Once time-variation and/or generated regressors are taken into account, neither Ramey's (2011b) defense news variable or the professional forecast of government spending Granger cause the Blanchard-Perotti shock.

First I'll test if Ramey's (2011b) defense news variable Granger causes government spending when time-varying parameters are allowed. Testing if the defense news variable Granger causes the Blanchard-Perotti shock can be done by testing if the defense news variable Granger causes government spending while controlling for the variables that generate the Blanchard-Perotti shock. Since the Granger causality test can be conducted on government spending directly, there is no need to conduct the test in two steps, thus eliminating potential generated regressor problems. The regression used for testing is

$$\frac{G_t - G_{t-1}}{G_{t-1}} = \sum_{j=1}^{p} \phi_{t,j} News_{t-j} + \gamma_t Controls_t + \varepsilon_t,$$

where the controls include some combination of lags of GDP growth, government spending growth, the effective tax rate, and the deficit as a percentage of GDP. A Granger causality test of the defense news Granger causing the Blanchard-Perotti shock can be written as $Pr(\phi_{t,j} = 0, \forall t, j)$. If $Pr(\phi_{t,j} = 0, \forall t, j) < .5$, then this is evidence in favor of the defense news variable Granger causing the Blanchard-Perotti shock. The closer $Pr(\phi_{t,j} = 0, \forall t, j)$ is to 0, the stronger the evidence of Granger causality. Details about how the time-varying parameter models are estimated are presented in the appendix. The tests use data from Ramey and Zubairy's (2018) data set. The effective sample period is from 1947-2015 and

1890-1946 is used for a training sample prior.[7]

Table 2.1 displays the results of the Granger Causality test probabilities for different sets of control variables.

baseline	baseline + tax and deficit	baseline + tax	baseline + deficit	joint
99.96%	100%	99.96%	99.99%	99.96%

Table 2.1: Granger Causality Test Probabilities (Defense News Variable)

The baseline model control variables are GDP growth and government spending growth. The test is conducted using different combinations of the baseline controls and a tax and deficit variable. The posterior probabilities of the defense news variable coefficients being zero, $Pr(\phi_{t,j} = 0, \forall t, j)$, is approximately 100% in each model. The joint probability across all of the models is 99.96%. The tests indicate that there is essentially no probability of the defense news variable Granger causing the Blanchard-Perotti shock.

Figure B.1 overlaps the posterior mean of the Blanchard-Perotti shocks from the time-varying parameter model including defense news and without defense news (the control variables include GDP growth, government spending growth, effective tax rate, and deficit as a percentage of GDP). As can be seen in the figure, the shocks are essentially identical, and the correlation between the two is approximately 98.6%. Figure B.2 overlaps the posterior mean of the Blanchard-Perotti shocks from time-invariant models including defense news and without defense news. When time-invariant models are used, there are more distinct differences, which is consistent with the results of Ramey (2011b). The correlation between the shocks drops to approximately 80%, and there are noticeable salient differences in the shocks, mostly around the late 40's early 50's. This indicates that the defense news variable is just picking up on structural breaks near the beginning of the sample, and since a time-invariant model does not allow for time variation, the model compensates by making the defense news variable coefficients significant, since the defense news variable is like a dummy

[7]The results are not sensitive to using 1939-1946 as a training sample prior.

variable.[8] This result is consistent with Perotti (2011) who shows that by simply dummying out just two quarters during WWII (when rationing was introduced) or during the Korean War (when new Federal regulations discouraging the purchase of durables was introduced) causes differences from using the defense news variable to disappear.

One may argue that the defense news variable is not as rich a variable as professional forecasts for government spending and that professional forecasts for government spending would Granger cause the Blanchard-Perotti shock. To test this, I take Auerbach and Gorodnichenko (2012) data and conduct the same test as I did for the defense news variable. The regression used for testing is

$$ln(G_t) = \sum_{j=1}^{p} \phi_{t,j} SPF_t + \gamma_t Controls_t + \varepsilon_t,$$

where controls include lags of log GDP, log government spending, log taxes. This is the baseline specification used in Auerbach and Gorodnichenko (2012). The Auerbach and Gorodnichenko (2012) data runs from 1966-2010. I use the first 10 years of data as a training sample and the effective sample is from 1976-2010. A Granger causality test of professional forecasts of government spending Granger causing government spending can be written as $Pr(\phi_{t,j} = 0, \forall t, j)$. Again, if $Pr(\phi_{t,j} = 0, \forall t, j) < .5$, then this is evidence in favor of the professional forecasts Granger causing government spending. In this case $Pr(\phi_{t,j} = 0, \forall t, j)$ is 94.63%.

Like Ramey (2011b), the results of Auerbach and Gorodnichenko (2012) do not hold in the time-varying parameter specification. There is essentially no evidence that professional forecasts of government spending Granger cause government spending, even in the time-invariant model. Further analysis indicates that just taking into account generated regressors eliminates the predictive power of the professional forecasts of government spending

[8] Approximately 70% of the defense news variable entries are zeros. Moreover, the defense news variable is a richer version of the Ramey and Shapiro (1998) war dates, which are dummy variables.

found in Auerbach and Gorodnichenko (2012). Figure B.3 overlaps the posterior mean of the Blanchard-Perotti shocks from time-varying parameter models including professional forecasts of government spending and without. As one can see in the figure, the shocks are essentially identical, and the correlation between the two is approximately 97%. When not allowing for time variation, the drop is actually pretty small, with the new correlation being approximately 94% (Figure B.4).

In summary, once taking into account time variation, I do not find evidence that the Ramey (2011b) defense news variable and the professional forecasts of government spending used in Auerbach and Gorodnichenko (2012) Granger cause the Blanchard-Perotti shock. The results hold even when I use forecast criteria such as mean squared error and mean absolute error, or when I take into account that the Granger Causality probabilities may be time-varying (i.e. that the expectations variables may be relevant during some periods but not others (Koop et al., 2010)). The results in this subsection are consistent with the theoretical argument made in the previous subsections, that once a more natural fiscal rule is used, the solution to fiscal foresight is not to use an expectations augmented model but to use a time-varying parameter model.

2.5 Estimating the Government Spending Multiplier

In this section I will discuss the estimation of government spending multipliers. First I will explain why time-varying parameter models should be preferred over regime-switching models when estimating fiscal multipliers. Second, I will discuss how to estimate multipliers using time-varying parameter Local Projections. Third, I will explain why expressing multipliers in terms of dollars may have biased estimates in previous studies and why the dollar based multiplier can be misleading.

2.6 Regime Dependence

In recent years, many have wondered if the size of the multiplier depends on the state of the economy. One argument for this is the Keynesian argument that when slack in the economy increases, increased government spending is less likely to crowd out private sector spending. Another argument, made by Christiano et al. (2011) and Woodford (2011) and others, is that when interest rates bind at the zero lower bound (ZLB), multipliers can be higher than normal. Hall (2009) and Barro and Redlick (2011) argue that the combined effect of price controls, rationing, the draft, and patriotism during WWII impacted the size of the multiplier (with Barro and Redlick (2011) arguing the effect to be positive and Hall (2009) believing the effect to be negative). The resurgence in this debate has led to many new empirical papers analyzing fiscal multipliers. Papers such as Auerbach and Gorodnichenko (2012, 2013); Riera-Crichton et al. (2015); Ramey and Zubairy (2018) and others have considered the possibility that multipliers may differ according to the state of the economy and estimate multipliers using regime-switching models.[9] What these papers do not take into account, however, is that fiscal multipliers may be changing in a continuous smooth fashion.

Outside of the fiscal foresight, there are several reasons to prefer time-varying parameter models when the goal is to test for state dependence. First, regime-switching models imply an unrealistic discontinuity in the response of the economy to a shock. For example, Ramey and Zubairy's (2018) baseline specification was a regime-switching model with a 6.5% unemployment rate threshold to estimate regime-switching multipliers. This modeling specification implies that the multiplier is significantly different for an economy with a 6.4% unemployment rate vs. a 6.5% unemployment rate despite those rates being practically the same. Second, regime-switching models assume that multipliers are constant within regimes; the model assumes that no matter how high the unemployment rate becomes (e.g. 10%),

[9] Multipliers estimated this way are often called state dependent multipliers. To avoid confusion with other uses of the word "state" in this paper, this paper will only refer to them as regime-switching multipliers.

the multiplier is the same as it would be at the threshold of 6.5%. Moreover theoretical models such as Michaillat (2014) imply that the government expenditure multiplier varies in a smooth continuous manner rather than in two discrete regimes.[10]

A third difficulty with regime-switching models is that they can have finite sample problems with some of their regimes, which can lead to imprecise multiplier estimates. For example, if one believes that multipliers differ during the business cycle, then it should not be far fetched to think that multipliers during the financial crisis of 2007-2009 may be significantly higher than multipliers during a typical recession.[11] This suggests that one would need to consider at least 3 regimes when measuring the multiplier. Since the recession regime already has few observations, and the financial crisis is a subset of the recession regime, it would be nearly impossible to obtain precise enough estimates of the multipliers to conduct meaningful inference.[12] This finite sample problem is not limited to that case. There may be several regimes that one wants to test (e.g. recession, slack, ZLB, uncertainty, debt to GDP, etc.), some of which may overlap. It may be the case that multipliers do not differ during recessions or the ZLB alone, but multipliers differ when both occur simultaneously. This may cause a finite sample problem because some regimes (e.g. recession regime and the ZLB) have so few observations. None of the empirical studies reviewed take into account the interaction between potential regimes, and not doing so may have biased their results.

Fourth, regime-switching models rely on the assumption that only the regime-switching

[10] Michaillat (2014) analyzes a public employment multiplier which is a type of government consumption multiplier. A public employment multiplier indicates the additional number of workers employed when one worker is hired in the public sector.

[11] Canzoneri et al. (2016) show that multipliers can be higher in recessions than in expansions if financial frictions are more sensitive to fiscal policy during recessions (a feature they find evidence for in their data). Hall (2013) found an unprecedented level of financial friction beginning in late 2008 that remained high four years later. None of the empirical papers reviewed take this into account multipliers during a financial crisis may differ from multipliers during a regular recession.

[12] Auerbach and Gorodnichenko (2012, 2013) solve this problem for expansion and recession regimes using regime-switching models where a conditional Bernoulli probability determines the regime. Financial crises are so unlikely, their methodology is unlikely to help. Furthermore, this methodology would not make sense for regimes such as the ZLB. This problem would also hold if slack regimes were used instead of recession regimes.

variable being tested can change the size of the multiplier. To see how this assumption could be problematic, take the following example. Say that the following model is estimated

$$\frac{Y_{t+h} - Y_{t-1}}{Y_{t-1}} = M_h(u_t)\frac{G_{t+h} - G_{t-1}}{Y_{t-1}} + \varepsilon_{t+h}^{(h)}, \tag{2.1}$$

where u_t is some measure of slack. M_h is the horizon h multiplier and is allowed to vary with the amount of slack. There are multiple reasons why the size of the multiplier may change that can be found in the literature (e.g., unemployment, ZLB, debt overhang, uncertainty, etc). Say the true model is

$$\frac{Y_{t+h} - Y_{t-1}}{Y_{t-1}} = M_h(u_t, ZLB_t, Debt_t, Uncertainty_t)\frac{G_{t+h} - G_{t-1}}{Y_{t-1}} + \varepsilon_{t+h}^{(h)}. \tag{2.2}$$

If the true model is equation 2, then estimation of equation 1 would lead to biased inference, since all changes (or lack thereof) in the multiplier would be attributed to the slack variable. Since one can never be sure how many things affect the size of the multiplier, it is sensible to just allow the multiplier to vary over time, not according to a specific variable.

A related problem with regime-switching models is that if regimes are not defined correctly, it will lead to invalid inference. For example, Ramey and Zubairy (2018) is the first study to model regime dependence involving the ZLB in the U.S. However as emphasized by Christiano et al. (2011) and Woodford (2011), multipliers may differ when the ZLB is binding for several periods. In other words, the multiplier may differ in the ZLB only when intermediate-maturity yields are constrained (Swanson and Williams, 2014), but Ramey and Zubairy (2018) define their ZLB regime as times when the short-term 3-month treasury yield remains relatively constant, which for their post WWII sample is from 2008Q4-2015Q4. According to the analysis in Swanson and Williams (2014), the ZLB is not necessarily a significant constraint when the short-term interest rate is near zero. In fact, they find that it is not until late 2011 that there is significant evidence of a substantially binding ZLB con-

straint. Again, theoretical models such as Christiano et al. (2011) and Woodford (2011) find that for a given path of fiscal stimulus, the longer the ZLB binds, the larger the multiplier. Since the length of time one expects the ZLB to bind varies over time, modeling the ZLB as a regime would be a misspecification because the multiplier should be allowed to vary with the expected duration of the ZLB constraint.

Lastly, unless explicitly modeled to do so, regime-switching models will not capture slowly changing structural instabilities.[13] For example, the increase in globalization may have decreased the size of the multiplier. Regime-switching models also cannot take into account changing institutional settings, changes in agents' expectations and uncertainty, or learning dynamics by private agents and/or the government. Modeling multipliers using time-varying parameter models can not only capture changing parameters over time but changing relationships between variables, thus avoiding all of the problems with regime-switching models listed up to this point. Finally, even if the true model is "regime switching", Granger (2008) showed that time-varying parameter models can approximate any non-linear model (non-linear in the variables and/or the parameters), which makes them more robust to model misspecification. This of course assumes that a sufficiently flexible time-varying parameter model is used and that the true time variation is not so violent that a time-varying parameter model cannot track the changes.[14]

[13] Auerbach and Gorodnichenko (2014) address this problem by using rolling window version of their regime-switching model of multipliers in Japan. This remedy creates finite sample issues due to the window size, which can be seen in the size of their confidence intervals.

[14] Baumeister and Peersman (2012) show via Monte Carlo simulations that time-varying parameter models are able to capture discrete breaks in a satisfactory manner should they occur.

2.7 Should We Express Multipliers in Terms of Dollars?

A baseline model used to represent multipliers is

$$\Delta ln(Y_{t+h}) = constant_{y,h,t} + \beta_{h,t} BPshock_t + \sum_k \psi_{y,h,k,t} z_{t-k} + \varepsilon^{(h)}_{y,t+h},\ h = 0, 1, 2, ...,$$

$$\Delta ln(G_{t+h}) = constant_{g,h,t} + \alpha_{h,t} BPshock_t + \sum_k \psi_{g,h,k,t} z_{t-k} + \varepsilon^{(h)}_{g,t+h},\ h = 0, 1, 2, ...,$$

Y is real GDP per capita and G is real government expenditure per capita.[15] The control vector z includes the growth rate of real GDP per capita, the growth rate of real government expenditure per capita, the change in the effective tax rate, and the change in the total deficit as a share of GDP (more details about the data can be found in the Appendix). $BPshock$ is the Blanchard-Perotti identified shock.[16,17] $\varepsilon_{y,t+h}$ and $\varepsilon_{g,t+h}$ are the horizon h time-varying local projection error terms for GDP and government spending respectively.

The time-varying cumulative multiplier for horizon h at time t is

$$M_{h,t} = \frac{\sum_{j=0}^{h} \beta_{j,t}}{\sum_{j=0}^{h} \alpha_{j,t}} \frac{Y_{t-1}}{G_{t-1}}.$$

The cumulative multiplier is the cumulative GDP responses divided by the cumulative government expenditure responses converted into its dollar equivalent and tells us the total h-period impact a spending shock has on GDP. For example, $M_{0,1947Q1}$ tells us the dollar impact an unexpected \$1 increase in spending at 1947Q1 has on GDP at 1947Q1. $M_{1,1947Q1}$ tells us the dollar impact an unexpected \$1 increase in spending at 1947Q1 has on GDP

[15] Government expenditures only includes government consumption and investment. Transfers are not included.

[16] The Blanchard-Perotti shock is generated by regressing $\frac{G_t - G_{t-1}}{G_{t-1}}$ on the vector of control variables; it is the forecast error from this regression. This is equivalent to a Cholesky decomposition in a VAR with government spending ordered first.

[17] The optimal lag length for the control vector is 1.

over the span of 1947Q1-1947Q2. $M_{h,1947Q1}$ tells us the dollar impact an unexpected \$1 increase in spending at 1947Q1 has on GDP over the span of 1947Q1-1947Q1+h. As argued by Mountford and Uhlig (2009); Uhlig (2010); Woodford (2011); Ramey (2016), cumulative multipliers address the relevant policy question because they take into account the total impact a government spending shock has on GDP for the relevant horizons.

I will now show that fiscal multipliers expressed in dollars are and will be time-varying. To understand why multipliers are time-varying it is helpful to decompose the cumulative multiplier into two components. Define the elastic multiplier which is just the cumulative multiplier before it is converted to its dollar equivalent

$$E_{h,t} = \frac{\sum_{j=0}^{h} \beta_{j,t}}{\sum_{j=0}^{h} \alpha_{j,t}}.$$

The cumulative multiplier is simply the elastic multiplier multiplied by the GDP to spending ratio

$$M_{h,t} = E_{h,t} \frac{Y_{t-1}}{G_{t-1}}.$$

Now lets assume that $\beta_{j,t}$ is constant for all t and $\alpha_{j,t}$ is constant for all t and j. This implies that the reactions of GDP and spending to spending shocks is constant over time and that the elastic multiplier static. But since the GDP to spending ratio $\frac{Y_{t-1}}{G_{t-1}}$ is not constant over time, the cumulative multiplier (in dollar terms) will be time-varying. This implies that the multipliers in studies that convert GDP and spending to the same units before estimation (i.e., Barro and Redlick (2011); Nakamura and Steinsson (2014); Ramey (2016); Barnichon and Matthes (2018b)) may be biased from the conversion since they do not estimate their models with time-varying parameter models. An alternative way to see why this is the case

is to look a the following model

$$\frac{Y_{t+h} - Y_{t-1}}{Y_{t-1}} = E_{h,t} \frac{G_{t+h} - G_{t-1}}{G_{t-1}} + \varepsilon_{t+h}^{(h)},$$

where $E_{h,t}$ is the elastic multiplier. Assuming that $E_{h,t}$ is constant over time, the model can be written as

$$\frac{Y_{t+h} - Y_{t-1}}{Y_{t-1}} = E_h \frac{G_{t+h} - G_{t-1}}{G_{t-1}} + \varepsilon_{t+h}^{(h)}. \tag{2.3}$$

Ignoring endogeneity issues (for simplicity), the elastic multiplier can be consistently estimated by OLS. If spending and GDP are converted to the same units before estimation then

$$\frac{Y_{t+h} - Y_{t-1}}{Y_{t-1}} = E_h \left(\frac{Y_{t-1}}{G_{t-1}} \right) \frac{G_{t+h} - G_{t-1}}{Y_{t-1}} + \varepsilon_{t+h}^{(h)}.$$

Reparameterization leads to

$$\frac{Y_{t+h} - Y_{t-1}}{Y_{t-1}} = M_{h,t} \frac{G_{t+h} - G_{t-1}}{Y_{t-1}} + \varepsilon_{t+h}^{(h)},$$

where $M_{h,t} = E_h \left(\frac{Y_{t-1}}{G_{t-1}} \right)$. Even though the elastic multiplier is constant and can be consistently estimated using OLS in equation 3, converting GDP and spending to the same units before estimation makes it necessary to use time-varying parameter models to estimate the cumulative multiplier. Using OLS or regime-switching models, as is done in the literature, leads to biased multiplier estimates since doing so would ignore the inherent time variation in the cumulative multiplier.

To illustrate, suppose for example that the contemporaneous elastic multiplier is constant over time and has a value of .2. If the GDP to spending ratio is constant over time and has a value of 5, then the dollar based multiplier would be constant over time and would

have a value of 1. In actuality, the GDP to spending ratio varies over time. The mean for the post WW2 sample period is approximately 5, but has values ranging from 4 to 6.5. This can be seen in the top portion of Figure B.5, which plots the GDP to spending ratio post 1947. Since the contemporaneous elastic multiplier is .2, the dollar based multiplier would vary from .8 to 1.3 over the sample period. This can be seen in the bottom portion of Figure B.5. Therefore converting spending and output to the same units before estimation can drastically impact inference since the multiplier would be some weighted average of the multipliers in Figure B.5. Dollar based multipliers below 1 indicate crowd-out, while dollar based multipliers above 1 indicates that an increase in government spending causes the private sector to spend more, so depending on the sample period used and how the time-varying multipliers are weighted, using the standard conversion techniques could lead to drastically different results. This simple example illustrates why the cumulative dollar based multiplier do not inform us if the impact of government spending changes over time.

Interestingly, if one were to assume that the cumulative dollar based multiplier is constant, this would imply that the elastic multiplier has to be time-varying. Assume that the contemporaneous dollar based multiplier is one, for example. This would imply that the contemporaneous elastic multiplier varies between .15 and .25 throughout the sample (see Figure B.6). Due to the implications of cumulative multiplier, I argue that the elastic multiplier should be used instead of the cumulative multiplier.

2.8 Concluding Remarks

This paper makes two major contributions. First, I show theoretically and empirically that fiscal foresight is not an omitted variable problem, and I explain why time-varying parameters models are a solution to the problem.

Second, I explain why not estimating fiscal multipliers using time-varying parameter models, and bit expressing fiscal multipliers in dollars can severely bias multiplier estimates and lead to incorrect inference.

Chapter 3

A Kalman Filter Test for Structural Instability

Researchers are often interested in whether there is parameter instability in regression models. Stock and Watson (1996); Ang and Bekaert (2002) show a significant amount of macroeconomic and financial time series exhibit structural instability.[1] Structural changes can occur for many reasons such as policy changes, technological evolution, changing economic conditions, etc. If structural instability is not appropriately taken into account, it can lead to invalid inference, poor out of sample forecasting, and incorrect policy evaluation. Since Chow's (1960) seminal work on structural instability at known break date(s), there have been many tests developed to detect instability at unknown dates and of unknown forms.[2] However, Cogley and Sargent (2005) demonstrated that the most popular frequentist tests for parameter instability such as the Nyblom-Hansen test (Nyblom, 1989; Hansen, 1992) and Quandt likelihood ratio (QLR) test (Quandt, 1960; Andrews, 1993) can have low power, even in moderately large samples. That is, even when there is parameter instability in mod-

[1] Parameter instability and structural instability will be used interchangeably. Variance instability and stochastic volatility will also be used interchangeably.

[2] See Hansen (2001) for a survey of the structural break literature in econometrics.

erately large samples, there is low probability that these tests will correctly reject the null hypothesis of parameter stability.

In order to take into account potential parameter instability, many studies have used Bayesian time-varying parameter models. Despite the seminal work of Cogley and Sargent (2005); Primiceri (2005) increasing the popularity of time-varying parameter models, few studies using these methods actually test for time variation (Chan, 2017). This is because testing would typically involve calculating the marginal likelihood, which is no trivial task for models estimated using MCMC. Even though there are numerous ways to calculate the marginal likelihood from MCMC draws (e.g., Gelfend and Dey (1994); Newton and Raftery (1994); Chib (1995); Meng and Wong (1996); Lewis and Raftery (1997); Chib and Jeliazkov (2001)), these methods can be computationally intensive or unstable.[3] Bayesian testing can also be sensitive to prior selection. This typically leads researchers to do sensitivity analysis, but even when non-informative priors are used, results can differ substantially.[4] One could just assume parameter instability and estimate a time-varying parameter model without testing, but if the true parameters are stable, this will lead to inefficient inference caused by overparameterization. Moreover, these models can be computationally demanding in higher dimensions due to the number of parameters being estimated, so assuming parameter instability does not come without costs.[5,6] This begets the need for a simple method to determine if parameters are time-varying.

I propose a Kalman filter test for parameter instability. The test can detect regression coefficient instability and variance instability. The instability can be from smooth continuous changes, abrupt discrete changes, or both. As shown in Granger (2008), time-varying pa-

[3]Primiceri (2005) does model comparison via Reversible Jump MCMC, but this too can be computationally demanding.

[4]See Phillips (1991a) and references therein for an example in the the unit root literature.

[5]To address the overparameterization problem, there is a growing literature on time-varying parameter shrinkage methods (e.g.,Bitto and Fruhwirth-Schnatter (2016); Nakajima and West (2012, 2013); Chan et al. (2016a)). These methods are also computationally burdensome.

[6]Computational burdens can be lessened to a degree using sampling algorithms from Chan and Jeliazkov (2009); McCauslanda et al. (2011).

rameter models can approximate any form of non-linearity which makes them more robust to model misspecification.[7,8] The Kalman filter test, which is formulated in a Bayesian model selection framework, uses a training sample prior and does not need sensitivity analysis by construction. As opposed to using cutoffs for binary reject or fail to reject decisions, the test calculates the posterior probability of time variation. The test does not require MCMC and is computationally simple.

When applied to a VAR of GDP, inflation, and a short-term interest rate, I find overwhelming evidence of time variation. The time variation comes in the form of stochastic volatility in all of the equations. When applied to a VAR of Smets and Wouters (2007) original data, I find overwhelming evidence of time variation. The time variation comes in the form of regression coefficient instability in every equation except the interest rate equation and stochastic volatility in every equation. Since it is standard in the DSGE literature to grade the fit and forecasting performance of the DSGE model relative to static VARs (Edge and Gurkaynak, 2010; Gurkaynak et al., 2013; Giannone et al., 2015), the results from the applications suggest that DSGE models should be judged on their performance relative to time-varying parameter VARs, not static VARs.

The paper is outlined as follows. The Kalman filter test is described in section 2. Section 3 describes an application of the test to a VAR of GDP, inflation, and a short-term interest rate. Section 4 describes an application of the test to a VAR of Smets and Wouters (2007) original data. Section 5 discusses extensions. Section 6 concludes.

Some notation: $N(\cdot,\cdot)$, $Beta(\cdot,\cdot)$, and $IG(\cdot,\cdot)$ are the normal, beta, and inverse-gamma distributions, respectively. $T_n(\cdot,\cdot)$ is the T-distribution with n degrees of freedom, and χ^2_n is the chi-squared distribution with n degrees of freedom. $y_{1:T} = \{y_1, ..., y_T\}$.

[7] This assumes that the parameter changes are not violent that a time-varying parameter model cannot track them.

[8] Baumeister and Peersman (2012) show via Monte Carlo simulations that time-varying parameter models are able to capture discrete breaks in a satisfactory manner should they occur.

3.1 The Kalman Filter Test for Structural Instability

3.1.1 Bayesian Time-Varying Parameter Overview

To understand the Kalman filter test, it is first necessary to give an overview of Bayesian time-varying parameter models. In this paper, a Bayesian time-varying parameter univariate regression is modeled as in West and Harrison (1997); Prado and West (2010):

$$y_t = X_t' \beta_t + \varepsilon_t \tag{3.1}$$

$$\beta_t = \beta_{t-1} + w_t \tag{3.2}$$

where β_t is the $p \times 1$ state vector at time t, X_t is a p-dimensional vector of regressors at time t, ε_t is the observation noise with $\varepsilon_t \sim N(0, v_t)$, w_t is the state evolution noise with $w_t \sim T_{n_{t-1}}(0, W_t)$, and $\varepsilon_t \perp w_t$. Unlike Cogley and Sargent (2005); Primiceri (2005) but like Cogley et al. (2010), the transition equation for the regression coefficients (equation 2) has a time-varying variance making the model less likely to perform poorly should an abrupt break occur (Koop and Potter, 2010).[9] Stochastic volatility is modeled as a Beta/Gamma random walk:

$$v_t = v_{t-1}/(\theta/\gamma_t) \tag{3.3}$$

where

$$\gamma_t \sim Beta\left(\frac{\theta n_{t-1}}{2}, \frac{(1-\theta)n_{t-1}}{2}\right),$$

[9] Baumeister and Peersman (2012, 2013) show via Monte Carlo simulations that time-varying parameter models are able to capture such discrete breaks in a satisfactory manner should they occur.

θ is a discount factor for stochastic volatility, and n_{t-1} is the degrees of freedom at time $t-1$. The model is estimated using discount factors and the Forward Filter Backward Sampler (FFBS) algorithm, and details about the estimation procedure can be found in the appendix.[10] Because discount factors and conjugate priors are used, MCMC is not needed.

Discount factors (also known as forgetting factors) are a natural framework for allowing and controlling for time variation in regression coefficients and the variance and are a core part of the Bayesian forecasting literature (West and Harrison, 1997; Prado and West, 2010). Discount factors lie in the interval $(0, 1]$. If a discount factor, say $\theta = .99$ is used, then from period $t \to t+1$, $\frac{1}{\theta} - 1 \approx 1\%$ of information known at time t is discounted or forgotten in the Kalman filtration process. And if $\theta = .99$, observations from 20 periods ago receive approximately 80% as much weight as this period's observation. If the discount factor is 1, no information is discounted and the model is static. The discounting of information over time allows more recent data to have a larger impact on the parameter value and is the crux for controlling for time variation in the parameters. The discount factors are estimated using Griddy Gibbs. Including the the discount factor as a parameter to be estimated takes into account uncertainty in the hyperparameters and is a natural way to safeguard against overfitting (Giannone et al., 2015).

The optimal the lag length is chosen by maximizing the joint log likelihood functions defined in terms of the predictive densities

$$\log[p(y_{1:T}|D_0, \delta, \theta, lag\, length)] = \sum_{t=1}^{T} \log[p(y_t|D_{t-1}, \delta, \theta, lag\, length)],$$

where

$$p(y_t|D_{t-1}, \delta, \theta, lag\, length),$$

[10] See Prado and West (2010); West and Harrison (1997) for derivations and more details about time-varying parameter models using conjugate priors and discount factors.

is the one step ahead predictive density, δ is the discount factor that controls for time variation in the regression coefficients, and D_{t-1} is the amount of information known at time $t-1$.[11] Maximizing the joint log likelihood functions is equivalent to maximizing the marginal likelihood. If each model is assumed to have the same prior probability, it is also equivalent to choosing the model with the highest posterior probability. Let $M_1, M_2, \ldots, M_I$ denote I models of the same structure that only differ in their lag lengths. The posterior probability for model i can be calculated by:

$$p(M_i|y_{1:T}, D_0) = \frac{p(M_i)p(y_{1:T}|D_0, M_i)}{\sum_{j=1}^{I} p(M_j)p(y_{1:T}|D_0, M_j)}.$$

Assuming all models have equal prior probability ($p(M_i) = I^{-1}\,\forall i$):

$$p(M_i|y_{1:T}, D_0) = \frac{p(y_{1:T}|D_0, M_i)}{\sum_{j=1}^{I} p(y_{1:T}|D_0, M_j)}.$$

Then conditional on the optimal lag length, the posterior distributions for the regression coefficients and the variance are model averaged over the grid of discount factors in order to take into account the uncertainty in the discount factors. Model averaging over the grid of discount factors is equivalent to placing a uniform prior on the discount factors and estimating them using Griddy Gibbs. Ideally one would use sampling importance resampling (see (Lopes et al., 1999)), but this is computationally impractical.

Discount factors are chosen by maximizing the joint log-likelihood functions defined in terms of the predictive densities:

$$log[p(y_{1:T}|D_0, \delta, \theta)] = \sum_{t=1}^{T} log[p(y_t|D_{t-1}, \delta, \theta)] \tag{3.4}$$

[11] X_t is suppressed in the marginal likelihood for clarity.

where

$$p(y_t|D_{t-1}, \delta, \theta) \tag{3.5}$$

is the one-step-ahead predictive density, D_{t-1} is the time $t-1$ information set, and δ is the discount factor that controls for time variation in the regression coefficients.[12] The one-step-ahead predictive densities are produced by the Kalman filter and are a measure of forecast performance. Maximizing the joint log-likelihood functions defined in terms of the predictive densities is equivalent to choosing the model with the highest marginal likelihood. If each model is assumed to have the same prior probability, it is also equivalent to choosing the model with the highest posterior probability. Let $M_1, M_2, \ldots, M_I$ denote I models of the same structure that only differ in their discount factors. The posterior probability for model i can be calculated by:

$$p(M_i|y_{1:T}, D_0) = \frac{p(M_i)p(y_{1:T}|D_0, M_i)}{\sum_{j=1}^{I} p(M_j)p(y_{1:T}|D_0, M_j)}.$$

In this paper, I use the common practice of assuming all models have equal prior probability ($p(M_i) = I^{-1}\,\forall i$). Therefore:

$$p(M_i|y_{1:T}, D_0) = \frac{p(y_{1:T}|D_0, M_i)}{\sum_{j=1}^{I} p(y_{1:T}|D_0, M_j)}.$$

The regression coefficients' discount factor is estimated over a default grid of [.7, 1] where the grid is partitioned by .01. The stochastic volatility discount factor is also chosen over a default grid of [.7, 1] where the grid is partitioned by .01. The initial grid size and partition are chosen because they cover fairly rapid parameter changes to no parameter change and should cover most situations (West and Harrison, 1997) .[13] It is important to note that if

[12] Lag length selection can also be incorporated if needed, which is done in the application sections.

[13] Depending on the context, these grid values may not be appropriate and can be adjusted accordingly. If desired, one can also conduct a sensitivity analysis with the size of the grid partitions.

posterior distribution of the discount factors pile up at the bottom of the grid, the grid must be lowered. For example let us say that the median regression coefficient discount factor is .95, but the median variance discount factor is .7. The grid for the variance discount factor must be lowered (e.g., to [.6, 1]). The reason for this is because the true discount factor for the variance may be .62 and the regression coefficient discount factor 1, but because the grid initially only searched over [.7, 1], it may be optimal for the regression coefficients to allow for time variation in order to compensate for the bound on the amount of stochastic volatility. Theoretically, one could allow just for regression coefficient instability or only for stochastic volatility. One would just have to restrict the discount factor not of interest to be equal to 1 and then search the grid for the other discount factor. This is not recommended because the restriction may exaggerate the results of the test. For example, if the true model has stochastic volatility and the test is restricted not to allow for stochastic volatility, it may be optimal for the time-varying parameter model to exaggerate the amount of time variation in the regression coefficients in order to compensate for the restriction.[14]

3.1.2 Prior

The default prior used for the Kalman filter test is a training sample prior similar to ones used in Cogley and Sargent (2005); Primiceri (2005); Cogley et al. (2010).[15] A training sample prior of size T_0 is:

$$\beta_0 | v_0 \sim N\left(\hat{\beta}_{OLS}, v_0 (X_0' X_0)^{-1}\right),$$

[14] A similar argument is made by Sims and Zha (2006) on an earlier version of Cogley and Sargent (2005) that did not allow for stochastic volatility in their time-varying parameter model.

[15] Cogley and Sargent (2005); Primiceri (2005); Cogley et al. (2010) use independent priors so their posterior distributions are only semi-conjugate and require MCMC for estimation. Primiceri (2005) made the priors weakly centered around the training sample prior while Cogley and Sargent (2005); Cogley et al. (2010) did not.

$$v_0 \sim IG\left(\frac{n_0}{2}, \frac{n_0 \hat{v}_{OLS}}{2}\right),$$

where n_0 is the prior degrees of freedom, $\hat{\beta}_{OLS}$ is the OLS estimate of the regression coefficients from the training sample, $\hat{v}_{OLS}$ is the OLS estimate of the variance from the training sample, and X_0 is the design matrix for the training sample. $n_0 = T_0 - p$, where p is the number of independent variables in the model, and T_0, the training sample size, is chosen by the researcher. It is typical for researchers to choose T_0 to be as small as possible in order to have as much data available for inference, but T_0 should be chosen so that n_0 is at least 5. Five degrees of freedom is the minimum because the marginal distribution for β_0 is a Student T distribution, and in order for the first four moments (and hence the mean, variance, skewness, and kurtosis) to be defined, the Student T distribution needs more than 4 degrees of freedom. Otherwise, the marginal prior would be flat and/or skewed, which could induce ill behaviors (Primiceri, 2005).[16,17] Because the training sample prior is used, it is recommended that the regressors and regressand be transformed so they do not contain a unit root.[18,19]

3.1.3 Kalman Filter Test

The test is conducted by first calculating the marginal likelihood of each model for each model in the [.6, 1]×[.6, 1] grid. Then the posterior probability of the static model is calculated. The probability that the model has time variation is the inverse of that, 1 minus

[16] If a flat prior is used and an influential outlier occurs early in the sample, it can adversely affect predictions which could exaggerate the amount of time variation in the model as the model attempt to explain the outlier and minimize the variance. Using a non-flat prior can stabilize inference when the sample size is low and outliers can be more influential (Hoff, 2009).

[17] It should be noted that non-informative priors (such as reference priors or $\beta_0|v_0$ having infinite variance) cannot be used in Bayesian model comparison due to Bartlett's paradox (Poirier, 1995).

[18] If variables contain unit roots, OLS results are biased in finite samples and can result in a large bias in sample sizes typically encountered in economic applications (Stock and Watson, 2007a) , thus making the training sample prior unreliable.

[19] If the researcher believes that the nonstationarity of a variable is caused by a deterministic trend, then one can just model the variable of interest around the trend and proceed using the training sample.

the probability of the static model. If it is optimal to discount, then it is evidence of time variation. If the posterior probability of the static model is 1, no information is discounted and one ends up with the static model. If the posterior probability of the static model is less than 1, then there is evidence of time variation. Obviously, the degree of the probability matters. If the posterior probability of the static model is 90%, the evidence of the static model is pretty high. If the model that maximizes the marginal likelihood has a variance discount factor less than 1, then it is evidence of structural instability in the variance. To calculate the amount of evidence, the model with the highest marginal likelihood is compared to the static model using posterior probabilities.

To illustrate how all of this works, here is an example. Let us say that I search over the grid to choose the optimal discount factors, and I find that .95 and .98 are the optimal discount factors for regression coefficients and the variance respectively. Let this model be denoted as M_1 (time-varying parameter model hypothesis). Let the static model, which has discount factors of 1, be denoted as M_2 (static model hypothesis). To calculate the amount of evidence for time variation, I calculate the posterior probability of M_1 vs M_2 which is:

$$p(M_1|y_{1:T}, X_{1:T}) = \frac{p(y_{1:T}|D_0, M_1)}{\sum_{j=1}^{2} p(y_{1:T}|D_0, M_j)}.$$

Theoretically, the test can be used to detect only for regression coefficient instability or only for stochastic volatility. One would just have to restrict the discount factor not of interest to be equal to 1 and then search the grid for the other discount factor. This is not recommended because the restriction may exaggerate the results of the test. For example, if the true model has stochastic volatility and the test is restricted not to allow for stochastic volatility, it may be optimal for the time-varying parameter model to exaggerate the amount of time variation in the regression coefficients in order to compensate for the restriction.[20]

[20] A similar argument is made by Sims (2001) on an earlier version of Cogley and Sargent (2005) that did not allow for stochastic volatility in their time-varying parameter model.

3.2 Application to GDP, Inflation, and a Short-Term Interest Rate

There have been many studies analyzing the structural instability in a model of GDP growth (e.g, McConnell and Perez-Quiros (2000); Stock and Watson (2002); Kim et al. (2004); Koop and Potter (2007, 2010)). Most of these studies find evidence of instability. Using a regime-switching model with two regimes, McConnell and Perez-Quiros (2000) find evidence of a structural break in the variance in the early 1980s. Using a regime-switching model with an unknown number of regime changes, Koop and Potter (2007) find substantial evidence of instability in the variance and modest at best evidence in the regression coefficients. Using a framework that approximately nests every popular model in the regime-switching and structural break literature Koop and Potter (2010), like Koop and Potter (2007), find substantial evidence of a decrease in stochastic volatility. Kim et al. (2004) investigate breaks in the volatility of various measures of aggregate activity. For most of the measures they consider, they find strong evidence of an abrupt break in the early 1980s. Stock and Watson (2002) find similar evidence for a change in volatility, but find the decline to have been more gradual.

There have been several studies about the instability in inflation. Cogley and Sargent (2005) present evidence of time variation in the inflation process both in the regression coefficients and variance. Stock (2001) finds little evidence for variation in the regression coefficients using classical methods, and Primiceri (2005) finds evidence for variation in the variance but little in the regression coefficients. Sims and Zha (2006) finds evidence of stochastic volatility but not regression coefficient instability. Cogley and Sargent (2005); Primiceri (2005); Sims and Zha (2006) also analyze potential instability in a short-term interest rate. Cogley and Sargent (2005) present evidence of time variation in the 3-month Treasury yield process in both the regression coefficients and variance. Primiceri (2005) finds some evidence

for variation in the variance in 3-month Treasury yield process but little in the regression coefficients. Sims and Zha (2006) find evidence of stochastic volatility in the federal funds rate process but not regression coefficient instability.

To estimate instability in the three variables, I run a trivariate VAR of GDP, inflation, and the 3-month Treasury yield from 1947Q3-2007Q4. [21] These variables are often used for small scale DSGE models (Clarida et al., 1999; Galí, 2015; An and Schorfheide, 2007) and are often of interest for forecasting. The three variable VAR is similar to the one used in Cogley and Sargent (2005); Primiceri (2005), but differ in a few ways. Instead of using the unemployment rate, I use real GDP per capita growth.[22] I use per capita growth instead of aggregate growth because population dynamics can affect parameter instability. Moreover, DSGE models are based on representative households, so using per capita GDP makes the most sense. I also use the PCE deflator to calculate inflation as opposed to the GDP deflator. The FOMC's preferred measure of inflation is PCE inflation (Bernanke, 2015). Lastly, like Cogley and Sargent (2005); Primiceri (2005), the three-month Treasury yield is used as the measure of short-term interest rate because it has a longer sample than the federal funds rate. Like Smets and Wouters (2007), the data in this study is expressed in quarterly as opposed to annual growth rates. All variables are transformed to be apparently stationary before estimation. In addition to using GDP growth per capita, inflation is first differenced, and the 3-month Treasury yield is first differenced. Inflation and the 3-month Treasury yield are first differenced to avoid issues in regards to unit roots. The training sample prior is based on the OLS results. If some of the variables contain unit roots, OLS results are biased in finite samples and the bias can be large in sample sizes typically encountered in economic applications (Stock and Watson, 2007a), which implies that the bias could be quite large for a training sample. If the training sample prior is not well calibrated, then it may be optimal

[21] The sample is ended in 2007 in order to avoid modeling issues with regards to the zero lower bound (ZLB). There is currently no consensus on how to deal with the ZLB, and it will not be addressed in this paper.

[22] GDP and the unemployment rate are considered interchangeable by Okun's Law (Stock and Watson, 2001).

for the model to allow for time variation in order to compensate for the prior, which could distort the results of the test. Transforming series to stationarity is often done (e.g., Koop and Korobilis (2009); Stock and Watson (2002, 2012)), and is a way to avoid modeling issues in regards to unit roots.

Lindé et al. (2016) highlighted current macroeconomic models performed poorly during the Great Recession. Considering the field often still uses static models, or do not allow for time variation in both the shock variances and structural parameters, it could be the case that the models performed so poorly in part because they do not take into account time variation in the parameters. Smets and Wouters (2007) compare their DSGE model to what they call a Bayesian VAR model (BVAR), which is a Bayesian VAR that has static regression coefficients, a static variance-covariance matrix, and uses a training sample prior. They demonstrate that the DSGE model outperforms the BVAR, which is one of the reasons why their model has become standard and prevalent (Edge and Gurkaynak, 2010).[23] Their main criterion used for comparison is the marginal likelihood, which is not only a measure of model fit but also a measure of out-of-sample forecasting performance (Smets and Wouters, 2007; Giannone et al., 2015). It is in fact pretty standard in the DSGE literature to validate a theoretical model by comparing its fit and forecasting performance to those of static VARs (Edge and Gurkaynak, 2010; Gurkaynak et al., 2013; Giannone et al., 2015). In Lindé et al. (2016), their augmented DSGE model is still being compared to a BVAR. Even though their models outperform the BVAR, the performance of DSGE models may be overstated if a time-varying parameter BVAR (TVP-BVAR) can significantly outperform a BVAR.

Smets and Wouters (2007) data set generally starts in 1947. They state that in previous versions of their paper they found that the first ten years are not representative of the rest of

[23] Smets and Wouters (2007) also compare their DSGE model to a Bayesian VAR that uses a Minnesota type prior from Sims and Zha (1998). The DSGE model outperforms the Sims and Zha (1998) model in terms of root mean squared error over multiple horizons. Smets and Wouters (2007) also conduct comparisons using the marginal likelihood, but since the Minnesota prior treats the variance-covariance matrix as known, the marginal likelihoods of the Sims and Zha (1998) model and the DSGE model should not be compared because the DSGE model full takes into account uncertainty while the Sims and Zha (1998) model does not.

the sample, so they decided to shorten the sample to 1957Q1 – 2004Q4.[24] The fact that they felt a stretch of data had to be omitted without any theoretical reasons indicates that their model may be misspecified because it does not factor in parameter instability. Periods such as the Great Recession are also not representative of their main sample and DSGE model forecasting breaks down during the Great Recession (Del Negro and Schorfheide, 2013; Lindé et al., 2016). In order to be able to use DSGE models for those periods, one cannot just omit the data because the model does not fit it well. The model should be augmented so it can be useful no matter the period. Smets and Wouters (2007) do attempt to investigate parameter instability by estimating their DSGE model over sub-periods and compare the estimates informally, but this would be similar to using a frequentist test for instability, which has been shown to have weak power relative to a time-varying parameter model (Cogley and Sargent, 2005). Smets and Wouters (2007) do not test for parameter instability in their BVARs.

To see if a TVP-BVAR would outperform a BVAR for a model of variables used in small scale DSGE systems, I conduct the Kalman filter test on each equation in the trivariate VAR. Table 3.1 presents the results of the Kalman filter test for the three variable VAR. 1947Q3-1955Q4 is used as a training sample making making the effective sample from 1956Q1-2007Q4. The table includes the optimal discount factors for each equation in the VAR. Lag length selection was also incorporated into the test. The optimal lag lengths were 1 for each equation.

	GDP	Inflation	3-Month T Bill
TVC	1	1	1
SV	.89	.82	.72

Note: TVC (time-varying coefficients) and SV (stochastic volatility). Effective sample period is 1956Q1-2007Q4.

Table 3.1: Kalman Filter Test Discount Factors for Small Scale VAR

I find evidence of stochastic volatility in all of the equations, while I do not find evidence

[24]Even though Smets and Wouters (2007) state that their sample starts in 1957, they include 1956 in their training sample prior. Because of this, I start the effective sample in 1956.

of regression coefficient instability in any of the equations. All of the stochastic volatility discount factors are appreciably less than 1, which indicates each equation has a vast amount of instability in the variance. I also calculate the posterior probability of time variation for each equation in the VAR. In each case, the time-varying parameter model hypothesis had a posterior probability of approximately 100%, providing overwhelming evidence against the static model. The results are consistent with Primiceri (2005). Primiceri (2005) computes posterior model probabilities of different hyperparameter values in the state equation for the time-varying regression coefficients. The model with the highest posterior probability was the one with the smallest prior variances in the state equation for the time-varying regression coefficients. The results are consistent with Sims and Zha (2006) who analyze similar data using regime-switching models and find that the model that performs the best allows for time variation in the variances only. The results are also consistent with Chan and Eisenstat (2017) who analyze similar data and find that constant coefficient VARs with stochastic volatility outperform both static VARs and TVP-VARs that allow for time variation in the regression coefficients and the variance.[25]

Cogley and Sargent (2005) showed that the most popular frequentist tests for parameter instability (specifically the Nyblom-Hansen test (Nyblom, 1989; Hansen, 1992), Andrews's (1993) sup-LM test, and Andrews's (1993) sup-Wald test) can have low power relative to their Bayesian time-varying parameter model.[26] Cogley and Sargent (2005) showed this by conducting a Monte Carlo simulation that used their time-varying parameter model to generate fictitious samples and then testing for time variation using the frequentist tests. It can be shown that the choice of prior shrinkage for the regression coefficients in Cogley and Sargent (2005) allows for variation in the regression coefficients roughly similar to that

[25]It should be noted that Primiceri (2005); Chan and Eisenstat (2017) test the entire system jointly, as opposed to equation by equation. Sims and Zha (2006) do their testing based on whether the equation is a policy equation or private sector equation.

[26]One may wonder if the type of instability specified in the alternative hypothesis for the frequentist test affects the power of the test. Elliott and Müller (2006) show that the most widely used structural break tests perform similarly, independently of the exact process of parameter instabilities.

allowed for by a regression coefficient discount factor of .99 (Koop and Korobilis, 2013). This implies that the Cogley and Sargent (2005) model can be thought of as a model in the model selection space of the Kalman filter test. This implies that their results from their Monte Carlo simulation would apply for the Kalman filter test.

Considering that it is popular to use frequentist tests for parameter instability, I feel that it may be of interest to see if the results of the Kalman filter test are in agreement with the results of the frequentist tests for the empirical examples. To compare the results to the most popular frequentist test, I conduct the Hansen (1992) and sequential LM tests (sup-LM, average Exponential LM, and the mean LM (Andrews and Ploberger, 1994)) on the three variable VAR over the same period.[27] For the Hansen (1992) test, only stochastic volatility tests are conducted. This is because the Hansen (1992) test cannot test for joint regression coefficient instability. I could use the Nyblom (1989) version which can test for instability in all regression coefficients jointly, but Monte Carlo simulations in Cogley and Sargent (2005) indicate that the sequential LM tests have similar if not more power, so I only use the Hansen (1992) test for stochastic volatility. Furthermore, as shown in Hansen (1992), when there is stochastic volatility, the test has trouble detecting regression coefficient instability.

Table 3.2 presents the stochastic volatility test results from the Hansen (1992) test.

GDP	Inflation	3-Month T Bill
1.647***	0.210	0.423*

Note: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.
Effective sample period is 1956Q1-2007Q4.

Table 3.2: Hansen Critical Values for Small Scale VAR

[27] The sequential LM tests are employed with a heteroskedasticity robust variance-covariance matrix.

The results provide evidence of stochastic volatility at the 1% level for the GDP equation and at the 10% level for the 3-month Treasury yield. The Hansen (1992) test fails to detect stochastic volatility at the 10% level for the inflation equation. This should be alarming considering that the Kalman filter test does not detect regression coefficient instability in the inflation equation but does detect a substantial amount of stochastic volatility, and the posterior probability of the time-varying parameter model hypothesis is approximately 100%. The order of magnitude of evidence between equations should also be alarming. The time-varying parameter hypothesis has approximately a 100% probability in each equation, but in terms of how much discounting is optimal for the stochastic volatility process, the 3-month Treasury yield equation requires the most with a discount factor of .72, the inflation equation comes in second with a discount factor of .82, and the GDP equation comes in third with a discount factor of .89. The Hansen (1992) test, however, finds the most evidence of stochastic volatility in the GDP equation which rejects at the 1% level, the second most in the 3-month Treasury yield which rejects at the 10% level, and the least in the inflation equation which fails to reject at the 10% level. Like the Kalman filter test, Cogley and Sargent (2005); Primiceri (2005); Sims and Zha (2006); Chan and Eisenstat (2017) also find considerable evidence of stochastic volatility.

Table 3.3 presents the results of the sequential LM tests. With the exception of the federal funds rate equation, the sequential LM tests and the Kalman filter tests are in agreement about the lack of evidence for regression coefficient instability. The Kalman filter test does not find evidence of regression coefficient instability in the federal funds rate equation. The sequential LM tests for the Treasury yield equation are in discord. The average exponential LM test finds evidence at the 10% level, while the sup LM and the mean LM tests fail to find evidence of instability at the 10% level. Recall that a regression coefficient discount factor of .99 yields time variation in the regression coefficients similar to Cogley and Sargent (2005). Since the frequentist tests have been shown to have low power relative to Cogley and Sargent's (2005) time-varying parameter model, and since the sup LM and mean LM

tests fail to reject at the 10% level, the significant result from the average exponential LM test is probably a false positive.

	GDP	Inflation	3-Month T Bill
Sup LM	0.976	0.498	0.108
Average Exponential LM	0.988	0.759	0.073*
Mean LM	0.987	0.824	0.109

Note: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Effective sample period is 1956Q1-2007Q4.

Table 3.3: Sequential LM Test P-Values for Small Scale VAR

It should be noted that Smets and Wouters (2007) use a medium scale 7 variable model, unlike the small scale 3 variable model used in this section. However, the overwhelming evidence of the time-varying parameter model dominating the static model should lead one to question how effective the standard New Keynesian DSGE model is in terms of performance versus time-varying parameter models.

3.3 Application to Smets and Wouters (2007) Data

As mentioned in the previous section, Smets and Wouters (2007) use a medium scale 7 variable model, unlike the small scale 3 variable model. It could be the case that the parameter instability found in the small scale model was due to omitted variable bias or that the instability occurred outside of Smets and Wouters (2007) original sample. So in this section, I apply the Kalman filter test to Smets and Wouters (2007) original data. The data set includes quarterly variables of GDP growth, consumption growth, investment growth, wage growth, log hours, inflation, and the federal funds rate. Even though it can be argued that log hours, inflation, and the federal funds rate are nonstationary, I do not transform the variables to apparent stationarity or make any alterations to Smets and Wouters (2007)

data in order to make things as comparable to their study as possible. 1947Q3-1955Q4 is used as a training sample making the effective sample from 1956Q1-2004Q4. The optimal lag lengths were 1 for each equation except for the federal funds rate equation (the optimal lag length was 2). The results of the Kalman filter test for the federal funds rate equation were not sensitive to whether I used 1 lag or 2, so the results for 1 lag are presented for all equations.

	GDP	Consumption	Investment	Wages	Hours	Inflation	Federal Funds Rate
TVC	.97	.97	.97	.98	.98	.95	1
SV	.91	.94	.92	.94	.95	.87	.78

Note: TVC (time-varying coefficients) and SV (stochastic volatility). Effective sample period is 1956Q1-2004Q4.

Table 3.4: Kalman Filter Test Discount Factors for Smets and Wouters VAR

The results of the Kalman filter test are presented in Table 3.4. As you can see, there is evidence of stochastic volatility for every equation . There is also evidence of regression coefficient instability for every equation except for the federal funds rate. The posterior probability of the time-varying parameter model hypothesis is approximately 100% for each equation, providing overwhelming evidence against the static model hypothesis.

I also conduct the Hansen (1992) and sequential LM tests on the data to compare the results. For the Hansen (1992) test for stochastic volatility (table 3.5), only the investment and wages equations fail to reject at the 10% level. Only the GDP, consumption and inflation equations reject at the 5% level. This should be alarming considering that the Kalman filter test detects stochastic volatility in every equation, and the posterior probability of the time-varying parameter model hypothesis is approximately 100% for each equation. Moreover, none of the discount factors for stochastic volatility are close to 1.

The results of the Hansen (1992) test were sensitive to lag length. When using 2 lags instead

GDP	Consumption	Investment	Wages	Hours	Inflation	Federal Funds Rate
1.092***	0.483**	0.322	0.268	.37*	0.632**	0.433*

Note: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.
Effective sample period is 1956Q1-2004Q4.

Table 3.5: Hansen Critical Values for Smets and Wouters VAR

of 1 for the federal funds rate equation, the test statistic goes from being significant at the 10% level to being significant at the 5% level (table 3.6).

Federal Funds Rate
0.480**

Note: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.
Effective sample period is 1956Q1-2004Q4.

Table 3.6: Hansen Critical Values for Interest Rate Equation with 2 Lags

The sequential LM tests (tables 3.7) are often not in agreement with each other, let alone with the Kalman filter test. Only 4 out of the 7 equations have at least one sequential LM test that rejects at the 10% level. Only 3 out of the 7 equations have at least one sequential LM test that rejects at the 5% level. For the GDP and consumption equations, the tests are not in agreement at the 10% level. For the consumption, inflation, and wages equations, the tests are not in agreement at the 5% level. The Kalman filter test and the sequential LM tests do agree on the lack of evidence or regression coefficient instability in the federal funds rate equation. The results of the sequential LM tests for the federal funds rate equation were not sensitive to whether I used 1 lag or 2.

One might argue the results from the Kalman filter test in the empirical applications are just false positives. Considering the sample sizes in the empirical applications and that the

	GDP	Consumption	Inv	Wages	Hours	Inflation	FF Rate
Sup LM	0.406	0.11	0.335	0.025**	0.482	0.001***	0.689
Ave Exp LM	0.241	0.054*	0.199	0.010***	0.369	0.001***	0.592
Mean LM	0.086*	0.026**	0.121	0.005***	0.234	0.023**	0.533

Note: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Effective sample period is 1956Q1-2004Q4.

Table 3.7: Sequential LM Test P-Values for Smets and Wouters VAR

posterior probability of time-varying parameter model hypothesis was approximately 100% for each equation, it is doubtful this is the case. Furthermore, frequentist tests have a size of 5% when using 5% significance levels (Stock and Watson, 2007a). Considering the evidence of the frequentist tests being in discord with one another at both the 5% and 10% level, it less likely that most of the significant frequentist results are false positives and much more likely that the the discord among the frequentist tests is due to the lack of power and that the significant results are the detection of actual time variation. Moreover, Elliott and Müller (2006) show that most frequentist structural break tests perform similarly, independently of the exact process of parameter instability in the alternative hypothesis. This implies that there would be no preferred test among the three, or among frequentist instability tests in general. Lastly, using similar data sets, Clark (2011); D'Agostino et al. (2013); Clark and Ravazzolo (2015); Aastveit et al. (2017) find that time-varying parameter models do better at real time forecasting for shorter and longer horizons than static models, implying that these systems probably have parameter instability.

3.4 Extensions

The test can be extended to detect regression coefficient instability for individual or a subset of regression coefficients as opposed to all regression coefficients jointly. To do so one would use block discounting (Prado and West, 2010). Using block discounting allows the researcher the flexibility to allow each regression coefficient or subsets of regression coefficients to have

different discount factors. However, it should be noted that as the number of discount factors becomes large, the computational demands increase exponentially because a grid must be searched for each discount factor.

The test can also be extended to multivariate systems. When combined with block discounting, one would have the flexibility of using one discount factor for the entire system, a different discount factor for each equation in the system, or different discount factors for each regression coefficient or subsets of regression coefficients. Multivariate systems can be estimated directly as a system (e.g., estimating the VAR jointly as opposed to equation by equation estimation) or through decoupling and recoupling where each univariate regression of the system is estimated and then recoupled into the multivariate system through a theoretical multivariate volatility structure defined by a sparse underlying graphical model (Zhao et al., 2016; Gruber and West, 2017).[28]

It may be possible to improve the test through Bayesian predictive synthesis. Bayesian predictive synthesis provides a theoretical basis for combining multiple forecast densities (McAlinn and West, 2018). Maximizing the joint log likelihood functions defined in terms of the one-step ahead predictive densities is equivalent to finding the model with the highest marginal likelihood. It could be the case however, that the model that is best for one-step ahead predictions is not best for multi-step ahead predictions. Combing the densities of multiple forecast horizons could decrease the potential of overfitting or giving too much weight to a spurious model. This may be of interest to those who want to test for instability in models used for forecasting at various horizons.

[28]See West and Harrison (1997); Prado and West (2010) for more details about estimating multivariate systems directly using discount factors and Zhao et al. (2016); Gruber and West (2017) for more details about estimation by decoupling and recoupling.

3.5 Conclusion

I have proposed a Kalman filter based test for parameter instability. The test can detect both regression coefficient instability and variance instability. The instability can be from smooth continuous changes, abrupt discrete changes, or both. The Kalman filter test, which is formulated in a Bayesian model selection framework, uses a training sample prior and does not need sensitivity analysis by construction. As opposed to using cutoffs for binary reject or fail to reject decisions, the test calculates the posterior probability of time variation. The test does not require MCMC and is computationally simple.

The test was illustrated via two applications: First, I applied the test to a vector autoregression (VAR) of variables often used in small scale dynamic stochastic general equilibrium (DSGE) models: GDP, inflation, and a short-term interest rate. Second, I applied the test to a VAR of Smets and Wouters (2007) original data. When applied to a VAR of GDP, inflation, and the 3-month Treasury yield, I find overwhelming evidence of time variation. The time variation comes in the form of stochastic volatility in all of the equations. When applied to a VAR of Smets and Wouters (2007) original data, I find overwhelming evidence of time variation. The time variation comes in the form of regression coefficient instability in every equation except the federal funds rate equation and stochastic volatility in every equation. The results from the two applications suggest that DSGE models be judged on their performance relative to time-varying parameter VARs instead of static VARs, since time-varying parameter models substantially outperform static Bayesian VARs for these data. Moreover, the results suggest that DSGE models should incorporate time variation into both their structural parameters and shock variances because linearized DSGE models, at least approximately, can be interpreted as restricted VARs (An and Schorfheide, 2007; Del Negro et al., 2007).[29] This would imply that since time-varying parameter VARs domi-

[29]DSGE models with time variation generally do not have a VAR representation (Graeve, 2017). Graeve (2017), however, shows that interpreting DSGE models with time variation in the parameters as a restricted time-varying parameter VARs typically holds.

nate static VARs for these systems, then DSGE models that allow for time variation, would dominate static DSGE models.

Bibliography

Aastveit, K. A., A. Carriero, T. E. Clark, and M. Marcellino (2017). Have standard vars remained stable since the crisis? *Journal of Applied Econometrics 32*(5), 931–951.

Alloza, M. (2015). Is fiscal policy more effective in uncertain times or during recessions? *Working Paper, University College London*.

An, S. and F. Schorfheide (2007). Bayesian analysis of dsge models. *Econometric Reviews 26*(2-4), 113–72.

Andrews, D. W., I. Lee, and W. Ploberger (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics 70*(1), 9–38.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*(6), 1383–1414.

Ang, A. and G. Bekaert (2002). Regime switches in interest rates. *Journal of Business and Economic Statistics 20*(2), 163–182.

Auerbach, A. J. and Y. Gorodnichenko (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy 4*(2), 1–27.

Auerbach, A. J. and Y. Gorodnichenko (2013). *Fiscal Multipliers in Recession and Expansion*, pp. 63–98. Fiscal Policy After the Financial Crisis.

Auerbach, A. J. and Y. Gorodnichenko (2014). Fiscal multipliers in japan. *NBER Working Paper 19911*.

Bachmann, R. and E. R. Sims (2012). Confidence and the transmission of government spending shocks. *Journal of Monetary Economics 59*(3), 235–249.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*(1), 47–78.

Baker, S. R. (2015). Debt and the consumption response to household income shocks. *Working Paper*.

Baker, S. R. (2018). Debt and the response to household income shocks: Validation and application of linked financial account data. *Journal of Political Economy 126*(4), 1504–1557.

Baker, S. R., N. Bloom, B. Canes-Wrone, S. J. Davis, and J. Rodden (2014). Why has us policy uncertainty risen since 1960? *American Economic Review 104*(5), 56–60.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics 131*(4), 1593–1636.

Barnichon, R. and C. Brownlees (2018). Impulse response estimation by smooth local projections. *Working Paper*.

Barnichon, R. and C. Matthes (2018a). Functional approximation of impulse responses. *Journal of Monetary Economics Forthcoming*.

Barnichon, R. and C. Matthes (2018b). Understanding the size of the government spending multiplier: It's in the sign. *Working Paper*.

Barro, R. J. and C. J. Redlick (2011). Macroeconomic effects from government purchases and taxes. *Quarterly Journal of Economics 126*(1), 51–102.

Barsky, R. B. and E. R. Sims (2012). Information, animal spirits, and the meaning of innovations in consumer confidence. *American Economic Review 102*(4), 1343–77.

Baumeister, C. and J. Hamilton (2018). Inference in structural vector autoregressions when the identifying assumptions are not fully believed: Re-evaluating the role of monetary policy in economic fluctuations. *Journal of Monetary Economics 100*, 48–65.

Baumeister, C. and G. Peersman (2012). The role of time-varying price elasticities in accounting for volatility changes in the crude oil market. *Journal of Applied Econometrics 28*(7), 1087–1109.

Baumeister, C. and G. Peersman (2013). Time-varying effects of oil supply shocks on the us economy. *American Economic Journal: Macroeconomics 5*(4), 1–28.

Beaudry, P. and F. Portier (2014). News-driven business cycles: Insights and challenges. *Journal of Economic Literature 52*(4), 993–1074.

Benhabib, J., B. Shapiro, and M. M. Spiegel (2018). How persistent are the effects of sentiment shocks? *Federal Reserve Bank of San Francisco Working Economic Letters 2018-22*.

Benhabib, J. and M. M. Spiegel (2018). Sentiments and economic activity: Evidence from u.s. states. *The Economic Journal Forthcoming*.

Berger, J. O. (2003). Could fisher, jeffreys and neyman have agreed on testing? *Statistical Science 18*(1), 1–32.

Berkowitz, J., I. Birgean, and L. Kilian (1999). On the finite-sample accuracy of nonparametric resampling algorithms foreconomic time series. In *Advances in Econometrics: Applying Kernel and Nonparametric Estimationto Economic Topics*, Volume 14, pp. 77–107.

Bernanke, B. (2015). The taylor rule: A benchmark for monetary policy?

Bhansali, R. J. (1997). Direct autoregressive predictors for multistep prediction: Order selection and performance relative to the plug in predictors. *Statistica Sinica 7*, 425–449.

Bitto, A. and S. Fruhwirth-Schnatter (2016). Achieving shrinkage in a time-varying parameter model framework. *Working Paper*.

Blanchard, O. (1993). Consumption and the recession of 1990-199. *American Economic Review 83*(2), 270–74.

Blanchard, O. and R. Perotti (2002). An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *Quarterly Journal of Economics 107*(4), 1329–1368.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica 77*(3), 623–685.

Bowman, D., D. Minehart, and M. Rabin (1999). Loss aversion in a consumption–savings model. *Journal of Economic Behavior & Organization 38*(2), 155–178.

Broda, C. and J. A. Parker (2014). The economic stimulus payments of 2008 and the aggregate demand for consumption. *Journal of Monetary Economics 68*(S), S20–S36.

Brodersen, K. H., F. Gallusser, J. Koehler, N. Remy, and S. L. Scott (2015). Inferring causal impact using bayesian structural time series models. *Annals of Applied Statistics 9*, 247–274.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B 37*(2), 149–192.

Brugnolini, L. (2018, August). About local projection impulse response function reliability. *CEIS Working Paper No. 440*.

Bunn, P., J. L. Roux, K. Reinold, and P. Surico (2017, February). The consumption response to positive and negative income shocks. *CEPR Discussion Paper 11829*.

Caldara, D. and C. Kamps (2017). The analytics of svars: A unified framework to measure fiscal multipliers. *Review of Economics Studies 84*(3), 1015–1040.

Canova, F. (2007). *Methods for Applied Macroeconomic Research*. Princeton, NJ: Princeton University Press.

Canova, F. and M. Ciccarelli (2009). Estimating multi-country var models. *International Economic Review 50*, 929–959.

Canzoneri, M., F. Collard, H. Dellas, and B. Diba (2016). Fiscal multipliers in recessions. *The Economic Journal 126*(590), 75–108.

Carvalho, C., E. Hsu, and F. Nechio (2016). Measuring the effect of the zero lower bound on monetary policy. *Federal Reserve Bank of San Francisco Working Paper 2016-06*.

Chan, J., E. Eisenstat, and R. Strachan (2016a). Stochastic model specification search for time-varying parameter vars. *Econometric Reviews 35*(8-10), 1638–1665.

Chan, J., E. Eisenstat, and R. Strachan (2016b). Stochastic model specification search for time-varying parameter vars. *Econometric Reviews 35*(8-10), 1638–1665.

Chan, J. C. (2017). Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews Forthcoming*.

Chan, J. C. and E. Eisenstat (2017). Bayesian model comparison for time-varying parameter vars with stochastic volatility. *Working Paper*.

Chan, J. C. and I. Jeliazkov (2009). Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation 1*(1), 101–120.

Chang, P.-L. and S. Sakata (2007). Estimation of impulse response functions using long autoregression. *The Econometrics Journal 10*(2), 453–469.

Chen, H., D. Levy, S. Ray, and M. Bergen (2008). Asymmetric price adjustment in the small. *Journal of Monetary Economics 55*, 728– 737.

Chib, S. (1993). Bayes regression with autoregressive errors. *Journal of Econometrics 58*, 275–294.

Chib, S. (1995). Marginal likelihood from the gibbs output. *Journal of the American Statistical Association 90*(432), 1313–1321.

Chib, S. (2001). *Handbook of Econometrics*, Volume 5, Chapter Markov Chain Monte Carlo Methods: Computation and Inference, pp. 3569–3649.

Chib, S. and I. Jeliazkov (2001). Marginal likelihood from the metropolis-hastings output. *Journal of the American Statistical Association 96*(493), 270–281.

Chib, S. and T. A. Kuffner (2016). Bayes factor consistency. *Working Paper*.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica 28*(3), 591–605.

Christelis, D., D. Georgarakos, T. Jappelli, L. Pistaferri, and M. van Rooij (2017, March). Asymmetric consumption effects of transitory income shocks. *DNB Working Paper No. 551*.

Christiano, L., M. Eichenbaum, and S. Rebelo (2011). When is the government spending multiplier large? *Journal of Political Economy 119*(1), 78–121.

Cimadomo, J., S. Hauptmeier, and M. Kirchner (2010). Transmission of fiscal shocks in the euro area: Time variation and driving forces. *Working Paper Series European Central Bank*.

Clarida, R., J. Galí, and M. Gertler (1999). The science of monetary policy: A new keynesian perspective. *Journal of Economic Literature 37*, 1661–1707.

Clark, T. (2011). Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics 29*(3), 327–341.

Clark, T. and F. Ravazzolo (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics 30*(4), 551–575.

Cloyne, J. and P. Surico (2017). Household debt and the dynamic effects of income tax changes. *Review of Economics Studies 84*(1), 45–81.

Cogley, T., G. Primiceri, and T. J. Sargent (2010). Inflation-gap persistence in the us. *American Economic Journal: Macroeconomics 2*(1), 43–69.

Cogley, T. and T. J. Sargent (2005). Drifts and volatilities: Monetary policies and outcomes in the post-ww ii us. *Review of Economic Dynamics. 8*(2), 262–302.

Coibion, O., Y. Gorodnichenko, and R. Kamdar (2018). The formation of expectations, inflation and the phillips curve. *Journal of Economic Literature 56*(4), 1447–1491.

D'Agostino, A., L. Gambetti, and D. Giannone (2011). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics 28*(1).

D'Agostino, A., L. Gambetti, and D. Giannone (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics 28*(1), 82–101.

Del Negro, M. and F. Schorfheide (2011). *Bayesian Macroeconometrics*, pp. 293–389. The Oxford Handbook of Bayesian Econometrics.

Del Negro, M. and F. Schorfheide (2013). *DSGE Model-Based Forecasting* (Handbook of Economic Forecasting ed.), Volume 2, pp. 57–140.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007). On the fit and forecasting performance of new-keynesian models. *Journal of Business and Economic Statistics 25*(2), 123–143.

Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold-mariano tests. *Journal of Business and Economic Statistics 33*(1).

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*.

Doms, M. and N. Morin (2004). Consumer sentiment, the economy, and the news media. *Federal Reserve Bank of St. Louis Working Paper 2004-09*.

Drautzburg, T. and H. Uhlig (2015). Fiscal stimulus and distortionary taxation. *Review of Economic Dynamics 18*(4), 894–920.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance 57*(1).

Edge, R. M. and R. S. Gurkaynak (2010). How useful are estimated dsge model forecasts for central bankers? *Brookings Papers on Economic Activity Fall*, 209–259.

Efron, B. and R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. CHAPMAN HALL/CRC.

Ellahie, A. and G. Ricco (2017). Government purchases reloaded: Informational insufficiency and heterogeneity in fiscal vars. *Journal of Monetary Economics 90*, 13–27.

Elliott, G. and U. K. Müller (2006). Efficient tests for general persistent time variation in regression coefficients. *Review of Economics Studies 73*(4), 907–940.

Fallick, B. C., M. Lettau, and W. L. Wascher (2016). Downward nominal wage rigidity in the united states during and after the great recession. *Finance and Economics Discussion Series 2016-001*.

Fernández-Villaverde, J., P. Guerrón-Quintana, K. Kuester, and J. Rubio-Ramírez (2015). Fiscal volatility shocks and economic activity. *American Economic Review 105*(11), 3352–3384.

Fernández-Villaverde, J., P. Guerrón-Quintana, and J. Rubio-Ramírez (2010). Fortune or virtue: Time-variant volatilities versus parameter drifting in u.s. data. *NBER Working Paper No. 15928*.

Fernández-Villaverde, J. and J. Rubio-Ramírez (2010). Macroeconomics and volatility: Data, models, and estimation. *NBER Working paper 16618*.

Galí, J. (2015). *Monetary Policy, Inflation, and Business Cycles*. Princeton, NJ: Princeton University Press.

Gelfend, A. and D. Dey (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society. Series B 56*(3), 501–514.

Gertler, M. and P. Karadi (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics 97*(2), 412–435.

Giannone, D., M. Lenza, and G. E. Primiceri (2018). Priors for the long run. *Journal of the American Statistical Association Forthcoming*.

Gillitzer, C. and N. Prasad (2018). The effect of consumer sentiment on consumption: Cross-sectional evidence from elections. *American Economic Journal: Macroeconomics 10*(4), 234–69.

Gonçalves, S. and H. White (2002). The bootstrap of themean for dependentheterogeneous arrays. *Econometric Theory 18*, 1367–1384.

Gonçalves, S. and H. White (2005). Bootstrap standard error estimates for linear regression. *Journal of the American Statistical Association 100*(471), 970–979.

Gorodnichenko, Y. (2014, July). Discussion of "government spending multipliers in good times and in bad: Evidence from u.s. historical data," by ramey and zubairy. *EFG meeting of the NBER Summer Institute*.

Graeve, F. D. (2017). A structural evaluation of vars with time-varying parameters and stochastic volatility. *Working Paper*.

Granger, C. and P. Newbold (1977). *Forecasting Economic Time Series*.

Granger, C. W. (2008). Non-linear models: Where do we go next - time varying parameter models? *Studies in Nonlinear Dynamics Econometrics 12*(3), Online.

Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Prentice Hall.

Gross, T., M. J. Notowidigdo, and J. Wang (2016). The marginal propensity to consume over the business cycle. *NBER Working paper 22518*.

Gruber, L. F. and M. West (2016). Gpu-accelerated bayesian learning and forecasting in simultaneous graphical dynamiclinear models. *Bayesian Analysis 11*(1), 125–149.

Gruber, L. F. and M. West (2017). Bayesian online variable selection and scalable multivariate volatility forecasting in simultaneous graphical dynamic linear models. *Econometrics and Statistics 3*, 3–22.

Guidolin, M. and D. L. Thornton (2011). Predictions of short-term rates and the expectations hypothesis. *Federal Reserve Bank of St. Louis Working Paper Series*.

Gurkaynak, R. S., B. Kisacikoglu, and B. Rossi (2013). Do dsge models forecast more accurately out-of-sample than var models? *Advances in Econometrics 32*, 27–80.

Hall, P., J. L. Horowitz, and B.-Y. Jing (1995). On blocking rules for the bootstrap with dependent data. *Biometrika 82*(3), 561–574.

Hall, R. E. (2009). By how much does gdp rise if the government buys more output? *Brookings Papers on Economic Activity Fall*, 183–231.

Hall, R. E. (2013). Financial frictions. *International Journal of Central Banking 9*(2), 155–163.

Hallin, M. (1986). Non-stationary q-dependent processes and time-varying moving-average models: Invertibility properties and the forecasting problem. *Advances in Applied Probability 18*(1), 170–210.

Hamilton, J. (1994). *Time Series Analysis*.

Hansen, B. E. (1990). Lagrange multiplier tests of parameter instability in non-linear model.

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of Policy Modeling 14*, 517–533.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in u.s. labor productivity. *Journal of Economic Perspectives 15*(4), 117–128.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46*(6), 1251–1271.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science 14*(4), 382–417.

Hoff, P. (2009). *A First Course in Bayesian Statistical Methods* (1 ed.). Springer Texts in Statistics. Springer-Verlag New York.

Hoff, P. and J. Wakefield (2013). Bayesian sandwich posteriors for pseudo-true parameters. *Journal of Statistical Planning and Inference 143*(10), 1638–1642.

Jappelli, T. and L. Pistaferri (2014). Fiscal policy and mpc heterogeneity. *American Economic Journal: Macroeconomics 6*(4), 107–36.

Jarocinski, M. and A. Marcet (2019). Priors about observables in vector autoregressions. *Journal of Econometrics Forthcoming*.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review 95*(1), 161–182.

Jordà, Ò. (2009). Simultaneous confidence regions for impulse responses. *Review of Economics and Statistics 91*(3), 629–647.

Jordà, Ò. and S. Kozicki (2011). Estimation and inference by the method of projection minimum distance: An application to the new keynesian hybrid phillips curve. *International Economic Review 52*(2), 461–487.

Jr, R. E. L. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy 1*, 19–46.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives 5*(1), 193–206.

Kaplan, G., B. Moll, , and G. L. Violante (2018). Monetary policy according to hank. *American Economic Review 108*(3), 697–743.

Kass, R. and A. Raftery (1995a). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kass, R. and L. Wasserman (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association 90*(431), 928–934.

Kass, R. E. and A. E. Raftery (1995b). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics 80*(2), 218–230.

Kilian, L. and Y. J. Kim (2011). How reliable are local projection estimators of impulse responses? *Review of Economics and Statistics 93*(4), 1460–1466.

Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Kim, C.-J., C. Nelson, and J. Piger (2004). The less-volatile u.s. economy: A bayesian investigation of timing, breadth, and potential explanations. *Journal of Business and Economic Statistics 22*(1), 80–93.

Koop, G. (1992). 'objective' bayesian unit root tests. *Journal of Applied Econometrics 7*(1), 65–82.

Koop, G. (2017). Bayesian methods for empirical macroeconomics with big data. *Review of Economic Analysis 9*(1), 33–56.

Koop, G. and D. Korobilis (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics 3*(4), 267–358.

Koop, G. and D. Korobilis (2013). Large time-varying parameter vars. *Journal of Econometrics 177*(2), 185–198.

Koop, G., D. Korobilis, and D. Pettenuzzo (2018). Bayesian compressed vector autoregressions. *Journal of Econometrics Forthcoming*.

Koop, G., R. Leon-Gonzalez, and R. W. Strachan (2010). Dynamic probabilities of restrictions in state space models: An application to the phillips curve. *Journal of Business and Economic Statistics 28*(3), 370–379.

Koop, G., H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics 74*(1).

Koop, G. and S. Potter (2007). Estimation and forecasting in models with multiple breaks. *Review of Economics Studies 74*(3), 763–789.

Koop, G. and S. Potter (2010). A flexible approach to parametric inference in nonlinear and time varying time series models. *Journal of Econometrics 159*(1), 134–150.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018). Har inference: Recommendations for practice. *Journal of Business and Economic Statistics 36*(4), 541–559.

Leduc, S. and Z. Liu (2016). Uncertainty shocks are aggregate demand shocks. *Journal of Monetary Economics 82*, 20–35.

Leeper, E. M., A. W. Richter, and T. B. Walker (2012). Quantitative effects of fiscal foresight. *American Economic Journal: Economic Policy 4*(2), 115–44.

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2008, May). Fiscal foresight: Analytics and econometrics. *NBER Working Paper No. 14028*.

Leeper, E. M., T. B. Walker, and S. S. Yang (2013). Fiscal foresight and information flows. *Econometrica 81*(3), 1115–1145.

Lewis, R. and G. C. Reinsel (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis 16*(3), 393–411.

Lewis, S. and A. Raftery (1997). Estimating bayes factors via posterior simulation with the laplace-metropolis estimator. *Journal of the American Statistical Association 92*, 648–655.

Leybourne, S. J., B. P. M. McCabe, and A. R. Tremayne (1996). Can economic time series be differenced to stationarity? *Journal of Business and Economic Statistics 14*(4), 435–446.

Liang, F., R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger (2008). Mixtures of g-priors for bayesian variable selection. *Journal of the American Statistical Association 103*(481), 410–423.

Lindé, J., F. Smets, and R. Wouters (2016, May). Challenges for central banks' macro models. *Sveriges Riksbank Working Paper Series*.

Ljungqvist, L. and T. J. Sargent (2012). *Recursive Macroeconomic Theory* (Fourth ed.).

Lopes, A. S. and M. C. Pereira (2014). Time-varying fiscal policy in the us. *Studies in Nonlinear Dynamics & Econometrics 18*(2), 157–184.

Lopes, H. F., A. R. B. Moreira, and A. M. Schmidt (1999). Hyperparameter estimationin forecast models. *Computational Statistics Data Analysis 29*, 387–410.

Lopes, H. F. and M. West (2004). Bayesian model assessment in factor analysis. *Statistica Sinica 14*, 41–67.

Lubik, T. A. and P. Surico (2010). The lucas critique and the stability of empirical models. *Journal of Applied Econometrics 25*(1), 177–194.

Ludvigson, S. C. (2004). Consumer confidence and consumer spending. *Journal of Economic Perspectives 18*(2), 29 –50.

Lusompa, A. (2019). Local projections, autocorrelation, and efficiency. *Working Paper*.

Mann, H. B. and A. Wald (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica 11*(3), 173–220.

Marcellino, M., J. Stock, and M. Watson (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of Econometrics 135*, 499–526.

McAlinn, K. and M. West (2018). Dynamic bayesian predictive synthesis in time series forecasting. *Journal of Econometrics Forthcoming*.

McCauslanda, W. J., S. Miller, and D. Pelletier (2011). Simulation smoothing for state–space models: A computational efficiency analysis. *Computational Statistics and Data Analysis 55*(1), 199–212.

McConnell, M. and G. Perez-Quiros (2000). Output fluctuations in the united states: What has changed since the early 1980s? *American Economic Review 90*(1464-76).

McCracken, M. and S. Ng (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business and Economic Statistics 34*(4), 574–589.

McCracken, M. W. and J. McGillicuddy (2018, June). An empirical investigation of direct and iterated multistep conditional forecasts. *Federal Reserve Bank of St. Louis Working Paper 2017-040B*.

Meng, X.-L. and W. Wong (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica 6*(4), 831–860.

Mian, A., K. Rao, and A. Sufi (2013). Household balance sheets, consumption, and the economic slump. *Quarterly Journal of Economics 128*(4), 1687–1726.

Michaillat, P. (2014). A theory of countercyclical government multiplier. *American Economic Journal: Macroeconomics 6*(1), 190–217.

Miranda-Agrippino, S. and G. Ricco (2018). The transmission of monetary policy shocks. *Working Paper*.

Misra, K. and P. Surico (2014). Consumption, income changes, and heterogeneity: Evidence from two fiscal stimulus programs. *American Economic Journal: Macroeconomics 6*(4), 84–106.

Mountford, A. and H. Uhlig (2009). What are the effects of fiscal policy shocks? *Journal of Applied Econometrics 24*(6), 960–992.

Müller, U. K. (2014). Hac corrections for strongly autocorrelated time series. *Journal of Business and Economic Statistics 32*(3), 311–322.

Murphy, K. M. and R. H. Topel (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics 3*(4), 370–379.

Nakajima, J. (2011). Time-varying parameter var model with stochastic volatility: An overview of methodology and empirical applications. *Monetary and Economic Studies 29*, 107–142.

Nakajima, J. and M. West (2012). Dynamic factor volatility modeling: A bayesian latent threshold approach. *Journal of Financial Econometrics 11*(1), 116–153.

Nakajima, J. and M. West (2013). Bayesian analysis of latent threshold dynamic models. *Journal of Business and Economic Statistics 31*(2), 151–164.

Nakamura, E. and J. Steinsson (2014). Fiscal stimulus in a monetary union: Evidence from us regions. *American Economic Review 104*(3), 753–792.

Nakamura, E. and J. Steinsson (2018a). High-frequency identification of monetary non-neutrality: The information effect. *Quarterly Journal of Economics 133*(3), 1283–1330.

Nakamura, E. and J. Steinsson (2018b). Identification in macroeconomics. *Journal of Economic Perspectives 32*(3), 59–86.

Newey, W. K. and D. McFadden (1994). *Large sample estimation and hypothesis testing*, Volume 4, Chapter 36, pp. 2111–2245. Amsterdam: Elsevier.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Newton, M. A. and A. E. Raftery (1994). Approximate bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society. Series B 56*(1), 3–48.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association 84*(405), 223–230.

Olea, J. L. M. and M. Plagborg-Møller (2018). Simultaneous confidence bands: Theory, implementation, and an application to svars. *Journal of Applied Econometrics Forthcoming*.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review 25*(1), 221–247.

Paul, P. (2018). The time-varying effect of monetary policy on asset prices. *Federal Reserve Bank of San Francisco Working Paper 2017-09*.

Paul, P. (Forthcoming). The time-varying effect of monetary policy on asset prices. *Review of Economics and Statistics*.

Peltzman, S. (2000). Prices rise faster than they fall. *Journal of Political Economy 108*(3), 466–502.

Perotti, R. (2011). Expectations and fiscal policy: An empirical investigation. *Working Paper*.

Perron, P. (2006). *Palgrave Handbook of Econometrics: Econometric Theory*, Volume 1, Chapter Dealing with structural breaks., pp. 278–352. Basingstoke, UK: Palgrave.

Pesaran, M. H., A. Pick, and A. Timmermann (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics 164*(1), 173–187.

Pesavento, E. and B. Rossi (2006). Small-sample confidence intervals for multivariate impulse response functions at long horizons. *Journal of Applied Econometrics 21*(8), 1135–1155.

Phillips, P. C. B. (1991a). Bayesian routes and unit roots: De rebus prioribus semper est disputandum. *Journal of Applied Econometrics 6*, 435–473.

Phillips, P. C. B. (1991b). To criticize the critics: An objective bayesian analysis of stochastic trends. *Journal of Applied Econometrics 6*, 333–364.

Plagborg-Møller, M. (2018). Bayesian inference on structural impulse response functions. *Quantitative Economics Forthcoming*.

Plagborg-Møller, M. and C. K. Wolf (2018). Local projections and vars estimate the same impulse responses. *Working Paper*.

Plagborg-Møller, M. and C. K. Wolf (2019). Local projections and vars estimate the same impulse responses. *Working Paper*.

Poirier, D. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach*. Cambridge: MIT Press.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*(428), 1303–1313.

Pope, A. L. (1990). Biases of estimators in multivariate non-gaussian autoregressions. *Journal of Time Series Analysis 11*(3), 249–258.

Poskitt, D. S. and W. Yao (2017). Vector autoregressions and macroeconomic modeling: An error taxonomy. *Journal of Business and Economic Statistics 35*(3), 407–419.

Prado, R. and M. West (2010). *Time Series: Modeling, Computation, and Inference*. Chapman Hall/CRC Press Taylor and Francis Group.

Priestley, M. B. (1988). *Non-linear and Non-stationary Time Series Analysis*. Academic Press.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economics Studies 72*(3), 821–852.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistic Association 55*(290), 324–330.

Raftery, A. (1995). Bayesian model selection in social research. *Sociological Methodology 25*, 111–163.

Ramey, V. A. (2009). Identifying government spending shocks: It's all in the timing. *NBER Working paper 15464*.

Ramey, V. A. (2011a). Can government purchases stimulate the economy? *Journal of Economic Literature 49*, 673–685.

Ramey, V. A. (2011b). Identifying government spending shocks: It's all in the timing. *Quarterly Journal of Economics 126*(1), 51–102.

Ramey, V. A. (2011c). A reply to roberto perotti's "expectations and fiscal policy: An empirical investigation". *Working Paper*.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2, Chapter 2, pp. 71–162.

Ramey, V. A. (2019). Ten years after the financial crisis: What have we learned from the renaissance in fiscal research? *Journal of Economic Perspectives 33*(2), 89–114.

Ramey, V. A. and M. D. Shapiro (1998). Costly capital reallocation and the effects of government spending. *Carnegie-Rochester Conference Series on Public Policy 48*(145-194).

Ramey, V. A. and S. Zubairy (2018). Government spending multipliers in good times and in bad: Evidence from u.s. historical data. *Journal of Political Economy 126*(2), 850–901.

Riera-Crichton, D., C. A. Vegh, and G. Vuletin (2015). Procyclical and countercyclical fiscal multipliers: Evidence from oecd countries. *Journal of International Money and Finance 52*(C), 15–31.

Romer, C. D. and D. H. Romer (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review 100*(3), 763–801.

Rossi, B. (2013). Advances in forecasting under instability. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting* (1 ed.), Volume 2B, pp. 1203–1324. Elsevier.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics 12*(4), 1151–1172.

Rudebusch, G. D. (2005). Assessing the lucas critique in monetary policy models. *Journal of Money, Credit and Banking 37*(2), 245–272.

Shapiro, A. H., M. Sudhof, and D. Wilson (2018). Measuring news sentiment. *Federal Reserve Bank of San Francisco Working Paper 2017-01*.

Shea, J. (1995a). Myopia, liquidity constraints, and aggregate consumption: A simple test. *Journal of Money, Credit, and Banking 27*(3), 798–805.

Shea, J. (1995b). Union contracts and the life-cycle/permanent-income hypothesis. *American Economic Review 85*(1), 186–200.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*(1), 1–48.

Sims, C. A. (1991). Comment by christopher a. sims on 'to criticize the critics'. *Journal of Applied Econometrics 6*(4), 423–434.

Sims, C. A. (2001). [evolving post-world war ii u.s. inflation dynamics]: Comment. *NBER Macroeconomics Annual 16*, 373–379.

Sims, C. A. (2002). The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity* (2), 1–40.

Sims, C. A. (2012). [disentangling the channels of the 2007–09 recession]: Comment. *Brookings Papers on Economic Activity Spring*, 141–148.

Sims, C. A., J. Stock, and M. Watson (1990). Inference in linear time series models with some unit roots. *Econometrica 58*(1), 113–144.

Sims, C. A. and T. Zha (1998). Bayesian methods for dynamic multivariate models. *International Economic Review 39*(4), 949–968.

Sims, C. A. and T. Zha (1999). Error bands for impulse responses. *Econometrica 67*(5), 1113–1155.

Sims, C. A. and T. Zha (2006). Were there regime switches in u.s. monetary policy? *American Economic Review 96*(1), 54–81.

Smets, F. and R. Wouters (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review 97*(3), 586–606.

Starr, M. (2012). Consumption, sentiment, and economic news. *Economic Inquiry 50*(4), 1097–1111.

Stock, J. (1994). *Unit Roots, Structural Breaks and Trends* (Handbook of Econometrics ed.), Volume 4, pp. 2740–2843. Amsterdam: Elsevier.

Stock, J. (2001). [evolving post-world war ii u.s. inflation dynamics]: Comment. *NBER Macroeconomics Annual 16*, 379–387.

Stock, J. and M. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics 14*(1), 11–30.

Stock, J. and M. Watson (2001). Vector autoregressions. *Journal of Economic Perspectives 15*(4), 101–115.

Stock, J. and M. Watson (2002). Has the business cycle changed and why? *NBER Macroeconomics Annual 17*.

Stock, J. and M. Watson (2007a). *Introduction to Econometrics*. Addison Wesley Longman.

Stock, J. and M. Watson (2007b). Why has u.s. inflation become harder to forecast? *Journal of Money Credit and Banking 39*, 3–33.

Stock, J. and M. Watson (2012). Disentangling the channels of the 2007-2009 recession. *Brookings Papers on Economic Activity Spring*, 81–135.

Stock, J. and M. Watson (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal 128*(610), 917–948.

Swanson, E. and J. Williams (2014). Measuring the effect of the zero lower bound on medium- and longer-term interest rates. *American Economic Review 104*, 3154–3185.

Uhlig, H. (1994). What macroeconomists should know about unit roots: A bayesian perspective. *Econometric Theory 10*(3/4), 645–671.

Uhlig, H. (2010). Some fiscal calculus. *American Economic Review 100*(2), 30–34.

Verdinelli, I. and L. Wasserman (1995). Computing bayes factors using a generalization of the savage-dickey density ratio. *Journal of the American Statistical Association 90*(430), 614–618.

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

White, H. (2001). *Asymptotic Theory for Econometricians*. Emerald Group Publishing Limited.

Woodford, M. (2011). Simple analytics of the government expenditure multiplier. *American Economic Journal: Macroeconomics 3*(1), 1–35.

Wu, J. C. and J. Zhang (2017). A shadow rate new keynesian model. *NBER Working Paper No. 22856*.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno De Finetti*, Volume 6, pp. 233–243. North-Holland, Amsterdam.

Zhao, Z. Y., M. Xie, and M. West (2016). Dynamic dependence networks: Financial time series forecasting and portfolio decisions. *Applied Stochastic Models in Business and Industry 32*(3), 311–332.

Appendix A

Appendix to Chapter 1

A.1 Normal Inverse-Wishart Posterior Equations

Let

$$y_t = B_1^{(1)} y_{t-1} + B_2^{(1)} y_{t-2} + \ldots + B_k^{(1)} y_{t-k} + u_t^{(0)},$$

as one would a standard Bayesian VAR. Define $\beta^{(0)} \equiv vec(\left[B_1^{(1)}, B_2^{(1)}, \ldots, B_k^{(1)}\right]')$, $X_t^{'(0)} \equiv I_n \otimes [y_{t-1}, y_{t-2}, \ldots, y_{t-k}]'$, then

$$y_t = X_t^{'(0)} \beta^{(0)} + u_t^{(0)},$$

where $u_t^{(0)} \sim N(0, \Sigma_u^{(0)})$. Assume a conditional normal inverse-Wishart prior for $p(\beta^{(0)}, \Sigma_u^{(0)})$. That is

$$p(\beta^{(0)} | \Sigma_u^{(0)}) \sim N(\underline{b}, \Sigma_u^{(0)} \otimes \underline{\Omega}),$$

$$p(\Sigma_u^{(0)}) \sim IW(\underline{n}, \underline{\Psi}),$$

where $\underline{b}$, $\underline{\Omega}$, $\underline{\Psi}$, and $\underline{n}$ are prior hyperparameters. Define $y \equiv [y'_{k+1}, \ldots, y'_T]'$ and $X \equiv [X_{k+1}^{(0)}, \ldots, X_T^{(0)}]'$, The posterior is also conditional normal inverse-Wishart. That is

$$p(\beta|\Sigma, y_{1:T}) \sim N(\bar{b}, \Sigma \otimes \overline{\Omega}),$$

$$p(\Sigma|y_{1:T}) \sim IW(\overline{n}, \overline{\Psi}),$$

where

$$\overline{\Omega} = (X'X + \underline{\Omega}^{-1})^{-1},$$

$$\hat{A} = (X'X)^{-1}X'y,$$

$$\overline{B} = \overline{\Omega}[\underline{\Omega}^{-1}\underline{B} + X'X\hat{A}],$$

$$\bar{b} = vec(\overline{B}),$$

$$\underline{b} = vec(\underline{B}),$$

$$S = (y - X\hat{A})'(y - X\hat{A}),$$

$$\overline{\Psi} = S + \underline{\Psi} + \hat{A}X'X\hat{A} + \underline{B}\underline{\Omega}^{-1}\underline{B} - \overline{B}'(X'X + \underline{\Omega}^{-1})\overline{B},$$

$$\overline{n} = \underline{n} + T - k.$$

A.2 Forward Filter Backward Sampler (FFBS)

A.2.1 Forward Filtering

More detail about the algorithm can be found in Prado and West (2010). Recall that a TVP model can be characterized as follows:

$$y_t = X_t'\beta_t + v_t,$$

$$\beta_t = \beta_{t-1} + w_t,$$

where y_t is a $r \times 1$ vector β_t is the $p \times 1$ state vector at time t, X_t is a $p \times r$ vector of regressors at time t, ϵ_t is a $r \times 1$ vector observation noise with $v_t \sim N(0, \Sigma_t)$, w_t is the state evolution noise with $w_t \sim N(0, \Sigma_t \otimes W_t)$, and v_s and w_t are independent and mutually independent $\forall s, t$. Notice that the variance of v_t is allowed to be time-varying. Stochastic volatility (time-varying variance) is modeled as a beta-Bartlett Wishart random walk. Stochastic volatility is modeled as a beta-Bartlett Wishart random walk which is defined as following $t-1$ to time t update

$$p(\Sigma_{t-1}|D_{t-1}) \sim IW(n_{t-1}, \Psi_{t-1})$$

then

$$p(\Sigma_t|D_{t-1}) \sim IW(\theta n_{t-1}, b_t\Psi_{t-1})$$

where θ is a discount factor for stochastic volatility and $b_t = (\theta n_{t-1}+k-1)/(n_{t-1}+k-1)$. Let D_0 represents initial prior information and the current information set represented by $D_t = \{D_{t-1}, y_t\}$. The estimates of a standard TVP DLM can be obtained as follows. First recall that for a VAR(k) $X_t \equiv I_n \otimes [y_{t-1}', \ldots, y_{t-k}']$. Imagine we have the posterior distributions of

β_t and v_t at time $t-1$. The posteriors are:

$$\beta_{t-1}|\Sigma_{t-1}, D_{t-1} \sim N(m_{t-1}, \Sigma_{t-1} \otimes C_{t-1}),$$

$$\Sigma_{t-1}|D_{t-1} \sim IW(n_{t-1}, \Psi_{t-1}),$$

where

$$M_t = M_{t-1} + A_t \epsilon_t^{'},$$

$$m_t = vec(M_t),$$

$$C_t = R_t - A_t A_t' q_t,$$

$$A_t = R_t X_t / q_t,$$

$$R_t = C_{t-1} + W_t = C_{t-1}/\delta.$$

$$n_t = \theta n_{t-1} + 1,$$

$$\Psi_t = \Psi_{t-1} + \epsilon_t \epsilon_t^{'} / q_t,$$

$$\epsilon_t = y_t - f_t,$$

$$f_t = X_t' M_{t-1},$$

$$q_t = X_t' R_t X_t' + 1,$$

where δ is the discount factor for the regression coefficients.The volatility evolves from the Σ_{t-1} posterior to the prior of Σ_t according to

$$p(\Sigma_t|D_{t-1}) \sim IW(\theta n_{t-1}, \theta \Psi_{t-1})$$

State evolves from the β_{t-1} prior to the β_t posterior as follows:

$$\beta_t|\Sigma_t, D_{t-1} \sim N(m_{t-1}, \Sigma_t \otimes R_t),$$

$\beta_t|D_{t-1}$ and $\Sigma_t|D_{t-1}$ are now the priors for β_t and Σ_t respectively. This leads to the following one-step-ahead predictive of y_t:

$$y_t|D_{t-1} \sim T_{\theta n_{t-1}}(f_t, q_t \frac{\Psi_{t-1}}{n_{t-1}}),$$

where The posterior for $\beta_t|D_t$ and $\Sigma_t|D_t$ can be now be calculated.

A.2.2 Backward Sampling

Initialize at T draw

$$\Sigma_T|D_T \sim IW(n_t, \Psi_t),$$

$$\beta_T|\Sigma_T, D_T \sim N(M_T, \Sigma_T \otimes C_T).$$

For $t-1$ to 1

$$\Sigma_t^{-1} = \theta\Sigma_{t+1}^{-1} + \gamma_t,$$

where

$$\gamma_t^{-1} \sim IW((1-\theta)n_t, \Psi_t),$$

and

$$\beta_t = m_t + \delta(\beta_{t+1} - m_t) + N(0, \Sigma_t \otimes C_t^*),$$

where

$$C_t^* = C_t - \delta^2 R_{t+1}.$$

A.3 Choosing Lag Length and Estimating Discount Factors

The optimal the lag length is chosen by maximizing the joint log likelihood functions defined in terms of the predictive densities

$$\log[p(y_{1:T}|D_0, \delta, \theta, lag\, length)] = \sum_{t=1}^{T} \log[p(y_t|D_{t-1}, \delta, \theta, lag\, length)],$$

where

$$p(y_t|D_{t-1}, \delta, \theta, lag\, length),$$

is the one step ahead predictive density, δ is the discount factor that controls for time variation in the regression coefficients, and D_{t-1} is the amount of information known at time $t-1$.[1] Maximizing the joint log likelihood functions is equivalent to maximizing the marginal likelihood. If each model is assumed to have the same prior probability, it is also equivalent to choosing the model with the highest posterior probability. Let $M_1, M_2, \ldots, M_I$ denote I models of the same structure that only differ in their lag lengths. The posterior probability

[1] X_t is suppressed in the marginal likelihood for clarity.

for model i can be calculated by:

$$p(M_i|y_{1:T}, D_0) = \frac{p(M_i)p(y_{1:T}|D_0, M_i)}{\sum_{j=1}^{I} p(M_j)p(y_{1:T}|D_0, M_j)}.$$

Assuming all models have equal prior probability ($p(M_i) = I^{-1}\,\forall i$):

$$p(M_i|y_{1:T}, D_0) = \frac{p(y_{1:T}|D_0, M_i)}{\sum_{j=1}^{I} p(y_{1:T}|D_0, M_j)}.$$

Then conditional on the optimal lag length, the posterior distributions for the regression coefficients and the variance are model averaged over the grid of discount factors in order to take into account the uncertainty in the discount factors. Model averaging over the grid of discount factors is equivalent to placing a uniform prior on the discount factors and estimating them using Griddy Gibbs. Ideally one would use sampling importance resampling (see (Lopes et al., 1999)), but this is computationally impractical.

The regression coefficients' discount factor is estimated over a default grid of [.7, 1] where the grid is partitioned by .01. The stochastic volatility discount factor is also chosen over a default grid of [.7, 1] where the grid is partitioned by .01. The initial grid size and partition are chosen because they cover fairly rapid parameter changes to no parameter change and should cover most situations (West and Harrison, 1997) .[2] It is important to note that if posterior distribution of the discount factors pile up at the bottom of the grid, the grid must be lowered. For example let us say that the median regression coefficient discount factor is .95, but the median variance discount factor is .7. The grid for the variance discount factor must be lowered (e.g., to [.6, 1]). The reason for this is because the true discount factor for the variance may be .62 and the regression coefficient discount factor 1, but because the grid initially only searched over [.7, 1], it may be optimal for the regression coefficients to allow for time variation in order to compensate for the bound on the amount of stochastic

[2]Depending on the context, these grid values may not be appropriate and can be adjusted accordingly. If desired, one can also conduct a sensitivity analysis with the size of the grid partitions.

volatility. Theoretically, one could allow just for regression coefficient instability or only for stochastic volatility. One would just have to restrict the discount factor not of interest to be equal to 1 and then search the grid for the other discount factor. This is not recommended because the restriction may exaggerate the results of the test. For example, if the true model has stochastic volatility and the test is restricted not to allow for stochastic volatility, it may be optimal for the time-varying parameter model to exaggerate the amount of time variation in the regression coefficients in order to compensate for the restriction.[3]

Depending on the situation more flexible time-varying parameter models may be needed. It is possible to allow subsets of regression coefficients to have different discount factors. To do so one, would use block discounting (Prado and West, 2010). However, it should be noted that as the number of discount factors becomes large, the computational demands increase exponentially because a grid must be searched for each discount factor. It is also possible to change discount factors over the sample period (Koop and Korobilis, 2013). Using cholesky style decoupling and recoupling (Zhao et al., 2016) or simultaneous graphical dynamic linear models(Gruber and West, 2016), it is also possible to allow each equation in a system to have different discount factors.

[3]A similar argument is made by Sims and Zha (2006) on an earlier version of Cogley and Sargent (2005) that did not allow for stochastic volatility in their time-varying parameter model.

A.4 Proofs of Consistency, Asymptotic Normality, and Efficiency of LP GLS

A.4.1 Preliminaries and Assumptions

Let y_t be an $r \times 1$ vector with Wold representation given by

$$y_t = \varepsilon_t + \sum_{h=1}^{\infty} \Theta_h \varepsilon_{t-h}$$

where ε_t is i.i.d. with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$ and the Θ_h satisfy $\sum_{h=0}^{\infty} \parallel \Theta_h \parallel < \infty$ where $\parallel \Theta_h \parallel^2 = tr(\Theta_h' \Theta_h)$ with $\Theta_0 = I_r$. Further, assume $det\{\Theta(z)\} \neq 0$ for $|z| \leq 1$ where $\Theta(z) = \sum_{h=0}^{\infty} \Theta_h z^h$ so that process can be written as an infinite order VAR representation

$$y_t = \sum_{j=1}^{\infty} A_j y_{t-j} + \varepsilon_t$$

with $\sum_{j=1}^{\infty} \parallel A_j \parallel < \infty$ and $A(z) = \Theta(z)^{-1}$. By recursive substitution

$$y_{t+h} = B_1^{(h)} y_t + B_2^{(h)} y_{t-1} + \ldots + \varepsilon_{t+h} + \Theta_1 \varepsilon_{t+h-1} + \ldots + \Theta_{h-1} \varepsilon_{t+1},$$

where $B_1^{(h)} = \Theta_h$, $B_j^{(h)} = \Theta_{h-1} A_j + B_{j+1}^{(h-1)}$ for $h \geq 1$ and with $B_{j+1}^{(0)} = 0$; $\Theta_0 = I_r$ with $j \geq 1$. The horizon h LP consists of estimating Θ_h from a least squares estimate of A_1^h with truncated regression

$$y_{t+h} = B_1^{(h)} y_t + \ldots + B_k^{(h)} y_{t-k+1} + e_{k,t+h}^{(h)},$$

where

$$e_{k,t+h}^{(h)} = \sum_{j=k+1}^{\infty} B_j^{(h)} y_{t-j+1} + \varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}.$$

For standard LP

$\hat{B}(k,h,OLS) = (\hat{B}_1^{(h)}, \ldots, \hat{B}_k^{(h)}) = \hat{\Gamma}'_{1-k,h}\hat{\Gamma}_k^{-1}$

$\hat{\Gamma}_{1-k,h} = (T-k-H)^{-1} \sum_{t=k}^{T-h} X_{t,k} y'_{t+h}$

$\hat{\Gamma}_k = (T-k-H)^{-1} \sum_{t=k}^{T-H} X_{t,k} X'_{t,k}$

$X_{t,k} = (y'_t, y'_{t-1,} \ldots, y'_{t-k+1})'$

$$\hat{B}(k,h,OLS) - B(k,h) = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{j=k+1}^{\infty} B_j^{(h)} y_{t-j+1} + \varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k}\} \hat{\Gamma}_k^{-1}$$

Define

$$U_{1T} = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{j=k+1}^{\infty} B_j^{(h)} y_{t-j+1}) X'_{t,k}\}$$

$$U_{2T} = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} \varepsilon_{t+h} X'_{t,k}\}$$

$$U_{3T} = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k}\}$$

A.4.2 Proof of Consistency for LP OLS Correction

Assumption 2. *Let y_t satisfy the Wold representation as presented above. Assume that in addition,*

$$(i)\ E|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}|< \infty$$

for $1 \leq i, j, k, l \leq n$.

(ii) k satisfies

$$\frac{k^2}{T} \to 0;\ T, k \to \infty$$

(iii) k satisfies

$$k^{1/2} \sum_{j=k+1}^{\infty} \parallel A_j \parallel \to 0\ T, k \to \infty.$$

These assumptions were used to show consistency of the VAR(∞) (Lewis and Reinsel, 1985) and the LP(∞) (Jordà and Kozicki, 2011).

Proposition 2. *Assume assumption 2 holds, then*

$$\parallel \hat{B}(k, h, OLS) - B(k, h) \parallel \xrightarrow{p} 0.$$

Proof. Lewis and Reinsel (1985) establish that $\parallel \hat{\Gamma}_k^{-1} \parallel_1$ is bounded in probability, so consistency in standard LP consists of showing that $\parallel U_{1T} \parallel, \parallel U_{2T} \parallel$, and $\parallel U_{3T} \parallel$ converge in probability to 0. This was shown in Jordà and Kozicki (2011). However, their proof showing $\parallel U_{3T} \parallel$ converging to 0 is incorrect. It is incorrect because $(\sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k}$ is assumed

to be independent across time. It is not. A correct proof is

$$U_{3T} = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k}\}$$

$$\| U_{3T} \|^2 = \| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k}\} \|^2$$

$$\| U_{3T} \|^2 = (T-k-H)^{-2} trace\{[\sum_{n=k}^{T-H} (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{n+h-l}) X'_{n,k}]'[\sum_{m=k}^{T-H} (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{m+h-l}) X'_{m,k}]\}$$

$$\| U_{3T} \|^2 = (T-k-H)^{-2} trace\{\sum_{m=k}^{T-H} \sum_{n=k}^{T-H} (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{n+h-l})' (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{m+h-l}) X'_{m,k} X_{n,k}\}$$

by the cyclic property of traces.

$$E \| U_{3T} \|^2 = (T-k-H)^{-2} trace \sum_{m=k}^{T-H} \sum_{n=k}^{T-H} E\{(\sum_{l=1}^{h-1} \Theta_l \varepsilon_{n+h-l})' (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{m+h-l}) X'_{m,k} X_{n,k}\}.$$

For $|n-m| > h-1$

$$E\{(\sum_{l=1}^{h-1} \Theta_l \varepsilon_{n+h-l})' (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{m+h-l}) X'_{m,k} X_{n,k}\}$$

$$= E\{(\sum_{l=1}^{h-1} \Theta_l \varepsilon_{n+h-l})'\} E\{(\sum_{l=1}^{h-1} \Theta_l \varepsilon_{m+h-l}) X'_{m,k} X_{n,k}\} = 0$$

by independence. So

$$E \| U_{3T} \|^2 = (T-k-H)^{-2} trace \sum_{m=k}^{T-H} \sum_{|n-m|<h} E\{(\sum_{l=1}^{h-1} \Theta_l \varepsilon_{n+h-l})' (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{m+h-l}) X'_{m,k} X_{n,k}\}.$$

Note that

$$| E\{(\sum_{l=1}^{h-1} \Theta_l \varepsilon_{n+h-l})' (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{m+h-l}) X'_{m,k} X_{n,k}\} |$$

$$\leq (E[\{(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})\}^2])^{1/2}(E[\{X'_{m,k}X_{n,k}\}^2])^{1/2}$$

by Cauchy-Schwarz inequality. And

$$X'_{m,k}X_{n,k} = y'_m y_n + y'_{m-1}y_{n-1} + \ldots + y'_{m-k+1}y_{n-k+1}$$

$$(X'_{m,k}X_{n,k})^2 = (y'_m y_n + y'_{m-1}y_{n-1} + \ldots + y'_{m-k+1}y_{n-k+1})^2$$

$$E[(X'_{m,k}X_{n,k})^2] = O_p(k^2)$$

and

$$\mid E([(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})]^2) \mid = constant$$

due to the finite fourth moments of ε and $\sum_{h=0}^{\infty} \parallel \Theta_h \parallel < \infty$. Consequently for $|n-m| \leq h-1$,

$$trace\{(E[\{(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})\}^2])^{1/2}(E[\{X'_{m,k}X_{n,k}\}^2])^{1/2}\} = O_p(k).$$

This implies there exists some finite constant M such that

$$E \parallel U_{3T} \parallel^2 = (T-k-H)^{-2} trace \sum_{m=k}^{T-H} \sum_{|n-m|<h} E\{(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l})'(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})X'_{m,k}X_{n,k}\}$$

$$\leq (T-k-H)^{-2}(T-k-H)(kh)M,$$

$$E \parallel U_{3T} \parallel^2 \leq (T-k-H)^{-1}k \times constant \xrightarrow{p} 0.$$

$$\Longrightarrow \parallel U_{3T} \parallel \xrightarrow{p} 0$$

That completes the correction that shows that LP OLS is consistent. □

A.4.3 Proof of Consistency for LP GLS

Theorem 2. *Assume assumption 2 holds, then for LP GLS*

$$\| \hat{B}(k,h,GLS) - B(k,h) \| \xrightarrow{p} 0.$$

Proof. To show consistency in LP GLS, there is an additional term

$$U_{4T} = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \hat{\Theta}_l \hat{\varepsilon}_{t+h-l}) X'_{t,k}\}$$

that must be taken into account. To see why note that the horizon h LP GLS is

$$y_{t+h} - \sum_{l=1}^{h-1} \hat{\Theta}_j \hat{\varepsilon}_{t+h-l} = B_1^{(h)} y_t + \ldots + B_k^{(h)} y_{t-k+1} + \widetilde{u}_{k,t+h}^{(h)},$$

where

$$\widetilde{u}_{k,t+h}^{(h)} = \sum_{j=k+1}^{\infty} B_j^{(h)} y_{t-j+1} + \varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l} - \sum_{l=1}^{h-1} \hat{\Theta}_j \hat{\varepsilon}_{t+h-l}.$$

and $\hat{\Theta}_h = \hat{B}_1^{(h)}$. To show consistency of LP GLS it suffices to show that $\| U_{4T} \| \xrightarrow{p} 0$ because for LP GLS

$$\| \hat{B}(k,h) - B(k,h) \| = \| U_{1T} \hat{\Gamma}_k^{-1} + U_{2T} \hat{\Gamma}_k^{-1} + U_{3T} \hat{\Gamma}_k^{-1} - U_{4T} \hat{\Gamma}_k^{-1} \|$$

$$\leq \| U_{1T} \| \| \hat{\Gamma}_k^{-1} \|_1 + \| U_{2T} \| \| \hat{\Gamma}_k^{-1} \|_1 + \| U_{3T} \| \| \hat{\Gamma}_k^{-1} \|_1 - \| U_{4T} \| \| \hat{\Gamma}_k^{-1} \|_1 .$$

Lewis and Reinsel (1985) establish that $\| \hat{\Gamma}_k^{-1} \|_1$ is bounded in probability. Jordà and Kozicki (2011) show $\| U_{1T} \|$ and $\| U_{2T} \|$ converges in probability to 0, and Proposition

2 shows $\| U_{3T} \|$ converges in probability to 0. The proof showing $\| U_{4T} \| \xrightarrow{p} 0$ will be a proof by induction. Assume the consistency for the previous $h-1$ horizons has been proven. Hence $\| \hat{\Theta}_l \| \xrightarrow{p} \| \Theta_l \| < \infty$ for $1 \leq l \leq h-1$. Note

$$\hat{\varepsilon}_t = \varepsilon_t + (\sum_{j=1}^{\infty} A_j y_{t-j}) - (\sum_{i=1}^{k} \hat{A}_i y_{t-i}).$$

Therefore

$$U_{4T} = \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \hat{\Theta}_l \hat{\varepsilon}_{t+h-l}) X'_{t,k}\}$$

$$= \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \hat{\Theta}_l (\varepsilon_{t+h-l} + (\sum_{j=1}^{\infty} A_j y_{t+h-l-j}) - (\sum_{i=1}^{k} \hat{A}_i y_{t+h-l-i}))) X'_{t,k}\}$$

$$= \sum_{l=1}^{h-1} \hat{\Theta}_l \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\varepsilon_{t+h-l} + (\sum_{j=1}^{\infty} A_j y_{t+h-l-j}) - (\sum_{i=1}^{k} \hat{A}_i y_{t+h-l-i}))) X'_{t,k}\}.$$

It was shown earlier that

$$\| U_{3T} \| = \| \sum_{l=1}^{h-1} \Theta_l \{(T-k-H)^{-1} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k}\} \| \xrightarrow{p} 0.$$

Since $h-1$ is finite and $\| \hat{\Theta}_l \| \xrightarrow{p} \| \Theta_l \| < \infty$

$$\| \sum_{l=1}^{h-1} \hat{\Theta}_l \{(T-k-H)^{-1} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k}\} \| \leq \sum_{l=1}^{h-1} \underbrace{\| \hat{\Theta}_l \|}_{bounded} \underbrace{\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k}\} \|}_{plim=0} \xrightarrow{p} 0.$$

To show $\| U_{4T} \| \xrightarrow{p} 0$ it suffices to show that

$$\| \sum_{l=1}^{h-1} \hat{\Theta}_l \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{j=1}^{\infty} A_j y_{t+h-l-j} - \sum_{i=1}^{k} \hat{A}_i y_{t+h-l-i}) X'_{t,k}\} \| \xrightarrow{p} 0.$$

Owing to $h-1$ in finite and $\| \hat{\Theta}_l \| \xrightarrow{p} \| \Theta_l \| < \infty$, this simplifies to showing

$$\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} (\sum_{j=1}^{\infty} A_j y_{t+h-l-j} - \sum_{i=1}^{k} \hat{A}_i y_{t+h-l-i}) X'_{t,k}\} \|$$

$$= \| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) - (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k}) X'_{t,k}\} \|$$

$$= \| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j})) X'_{t,k}\} -$$

$$\{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k}) X'_{t,k}\} \| \xrightarrow{p} 0.$$

Jordà and Kozicki (2011) and Lewis and Reinsel (1985) already showed

$$\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j})) X'_{t,k}\} \| \xrightarrow{p} 0.$$

Now all that is left to show is

$$\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k}) X'_{t,k}\} \| \xrightarrow{p} 0.$$

Note that $(\hat{B}(k,1) - B(k,1)$ does not depend on the t subscript so it can be factored out. That is,

$$\| \{(T-k-H)^{-1} \sum_{t=k}^{T-H} ((\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k}) X'_{t,k}\} \|$$

$$= \| \{(\hat{B}(k,1) - B(k,1))(T-k-H)^{-1} \sum_{t=k}^{T-H} X_{t+h-l-1,k} X'_{t,k}\} \|$$

$$\leq \underbrace{\| \{(\hat{B}(k,1) - B(k,1)) \|}_{plim=0} \underbrace{\| (T-k-H)^{-1} \sum_{t=k}^{T-H} X_{t+h-l-1,k} X'_{t,k}\} \|_1}_{bounded} \xrightarrow{p} 0.$$

Since this is a proof by induction, it was assumed that the first $h-1$ horizons are consistent, so the first term converges in probability. The second term is bounded due to $\| \hat{\Gamma}_k \|_1 = \| (T-k-H)^{-1} \sum_{t=k}^{T-H} X_{t,k} X'_{t,k} \|_1$ being bounded and since the autocovariances are absolutely summable. It follows that

$$\| \hat{\Theta}_l \{ (T-k-H)^{-1} \sum_{t=k}^{T-H} ((\varepsilon_{t+h-l} + (\sum_{j=1}^{\infty} A_j y_{t+h-l-j}) - (\sum_{i=1}^{k} \hat{A}_i y_{t+h-l-i}))) X'_{t,k} \} \| \xrightarrow{p} 0$$

for each $1 \leq l \leq h-1$. Therefore, $\| U_{4T} \| \xrightarrow{p} 0$ since the sum of a finite number of terms that each converge to zero also converges to 0. That is

$$\| U_{4T} \| = \| \sum_{l=1}^{h-1} \hat{\Theta}_l \{ (T-k-H)^{-1} \sum_{t=k}^{T-H} ((\varepsilon_{t+h-l} + (\sum_{j=1}^{\infty} A_j y_{t+h-l-j}) - (\sum_{i=1}^{k} \hat{A}_i y_{t+h-l-i}))) X'_{t,k} \} \| \xrightarrow{p} 0.$$

To complete the proof by induction, note that the horizon 0 LP is a VAR, and the consistency results for the VAR were proved in Lewis and Reinsel (1985), so the first step in the induction process was proved. □

A.4.4 Proof of Asymptotic Normality for LP OLS Correction

Assumption 3. *Let y_t satisfy the Wold representation as presented in the preliminary section. Assume that in addition,*

(i) $E|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}| < \infty$

for $1 \leq i, j, k, l \leq r$.

(ii) k satisfies

$$\frac{k^3}{T} \to 0; \; T, k \to \infty$$

(iii) k satisfies

$$T^{1/2} \sum_{j=k+1}^{\infty} \parallel A_j \parallel \rightarrow 0 \; T, k \rightarrow \infty.$$

Proposition 3. *Assume assumption 3 holds, then for LP OLS*

$$\sqrt{T-k-H} vec[\hat{B}(k,h,OLS) - B(k,h)] \xrightarrow{d} N(0, \Omega_h).$$

where

$$\Omega_h = \sum_{t=0}^{\infty} \sum_{s=0}^{\infty} cov[vec\{(\varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k} \Gamma_k^{-1}\}, vec\{(\varepsilon_{s+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{s+h-l}) X'_{s,k} \Gamma_k^{-1}\}'].$$

These assumptions were used to show asymptotic normality of the VAR(∞) (Lewis and Reinsel, 1985) and the LP(∞) (Jordà and Kozicki, 2011). It turns out Jordà and Kozicki (2011) use the incorrect Central Limit Theorem. Jordà and Kozicki (2011) proof follows the same argument as Lewis and Reinsel (1985). Lewis and Reinsel (1985) use a martingale CLT to prove asymptotic normality. This is possible because in the case of a VAR since

$$vec\{(T-k-H)^{-1/2} \sum_{t=k}^{T-H} (\varepsilon_{t+1} X'_{t,k}) \Gamma_k^{-1}\}$$

is a martingale, because ε_{t+1} and $X'_{t,k}$ are independent of each other, and ε_{t+1} is an i.i.d. and is therefore uncorrelated over time. In order to use the martingale CLT theorem for standard LP

$$vec\{(T-k-H)^{-1/2} \sum_{t=k}^{T-H} (\varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k} \Gamma_k^{-1}\}$$

would need to be a martingale. But it is not a martingale. Even though $(\varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l})$ is independent of $X'_{t,k}$, the process is not a martingale because the error term $(\varepsilon_{t+h} +$

$\sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l})$ is correlated across h horizons and $X'_{t,k}$ is correlated for potentially infinite horizons. Instead of using the Martingale Central Limit Theorem, the Gordin Central Limit Theorem should have been used. Given that the ε_t are i.i.d. and strongly stationary, $(\varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l})$ are strongly stationary and ergodic. Due to the assumptions placed on y_t, $X'_{t,k}$ is strongly stationary and ergodic. Hence

$$\{vec\{(\varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k} \Gamma_k^{-1}\}\}_{t=-\infty}^{t=\infty}$$

is strongly stationary and ergodic (Hayashi, 2000; White, 2001). The Gordin CLT states that if a time series process is strongly stationary and ergodic and satisfies the following three conditions:

1. Asymptotic uncorrelatedness
2. Summability of autocovariances
3. Asymptotic negligibility of innovations,

then it is asymptotically normal (Greene, 2012). The corrected proof of standard LP can be shown as follows.

Proof. To show asymptotic uncorrelatedness need to show that

$$lim_{j\to\infty} E[\{(\varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k}\} \Gamma_k^{-1} \Big| \{(\varepsilon_{t+h-j} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l-j}) X'_{t-j,k}\} \Gamma_k^{-1}] = 0,$$

where $E[\cdot \Big| \cdot]$ is the conditional expectation. Asymptotic uncorrelatedness is trivially satisfied because when j is greater than $h-1$, the process is independent since $(\varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l})$ would be independent of $(\varepsilon_{t+h-j} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l-j})$.

To show Summability of autocovariances, need to show

$$lim_{T\to\infty} var((T-k-H)^{-1/2} vec\{\sum_{t=k}^{T-H}(\varepsilon_{t+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{t+h-l}))X'_{t,k}\Gamma_k^{-1}\})$$

is finite and constant. Define

$$s_T = (T-k-H)^{-1/2} vec\{\sum_{t=k}^{T-H}(\varepsilon_{t+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{t+h-l}))X'_{t,k}\Gamma_k^{-1}\}.$$

Note that

$$vec\{(\varepsilon_{t+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{t+h-l})X'_{t,k}\Gamma_k^{-1}\} = (\Gamma_k^{-1}X_{t,k}\otimes I_r)vec((\varepsilon_{t+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{t+h-l}))$$

so

$$var(s_T) = (T-k-H)\sum_{m=k}^{T-H}\sum_{n=k}^{T-H} E[(\Gamma_k^{-1}X_{m,k}\otimes I_r)vec((\varepsilon_{m+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l}))$$

$$\times vec((\varepsilon_{n+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l}))'(X'_{n,k}{\Gamma'_k}^{-1}\otimes I_r)]$$

for $|n-m|> h-1$ the most future $(\varepsilon_{t+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{t+h-l})$ in the couple is independent of everything else. Therefore

$$var(s_T) = (T-k-H)^{-1}\sum_{m=k}^{T-H}\sum_{|n-m|<h} E[(\Gamma_k^{-1}X_{m,k}\otimes I_r)vec((\varepsilon_{m+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l}))$$

$$\times vec((\varepsilon_{n+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l}))'(X'_{n,k}{\Gamma'_k}^{-1}\otimes I_r)].$$

If one conditions on information known up to time n ($\mathcal{F}_n$ will denote the time n information

set)

$$E[(\Gamma_k^{-1}X_{m,k}\otimes I_r)vec((\varepsilon_{m+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l}))vec((\varepsilon_{n+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{n+h-l}))'(X'_{n,k}\Gamma'^{-1}_k\otimes I_r)\mid\mathcal{F}_n]$$

$$=[(\Gamma_k^{-1}X_{m,k}\otimes I_r)\Sigma_{e^{(h)},(m-n)}(X'_{n,k}\Gamma'^{-1}_k\otimes I_r)]$$

where

$$\Sigma_{e^{(h)},(m-n)}=E[(\varepsilon_{m+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})(\varepsilon'_{n+h}+\sum_{l=1}^{h-1}\varepsilon'_{n+h-l}\Theta'_l)\mid\mathcal{F}_n]$$

$$=E[(\varepsilon_{m+h}+\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})(\varepsilon'_{n+h}+\sum_{l=1}^{h-1}\varepsilon'_{n+h-l}\Theta'_l)]$$

which is constant and finite for all m and n due to the finite fourth moments of ε and $\sum_{h=0}^{\infty}\parallel\Theta_h\parallel<\infty$.

$$[(\Gamma_k^{-1}X_{m,k}\otimes I_r)\Sigma_{e^{(h)},(m-n)}(X'_{n,k}\Gamma'^{-1}_k\otimes I_r)]$$

$$=[(\Gamma_k^{-1}X_{m,k}\otimes I_r)(1\otimes\Sigma_{e^{(h)},(m-n)})(X'_{n,k}\Gamma'^{-1}_k\otimes I_r)]$$

$$=[(\Gamma_k^{-1}X_{m,k}\otimes\Sigma_{e^{(h)},(m-n)})(X'_{n,k}\Gamma'^{-1}_k\otimes I_r)]$$

$$=[\Gamma_k^{-1}X_{m,k}X'_{n,k}\Gamma'^{-1}_k\otimes\Sigma_{e,m,n}]$$

and

$$E[\Gamma_k^{-1}X_{m,k}X'_{n,k}\Gamma'^{-1}_k\otimes\Sigma_{e^{(h)},(m-n)}]=\Gamma_k^{-1}\Gamma_{(m-n),k}\Gamma'^{-1}_k\otimes\Sigma_{e^{(h)},(m-n)}$$

where $E(X_{m,k}X'_{n,k})=\Gamma_{(m-n),k}$. Due to

$$E[\Gamma_k^{-1}X_{m,k}X'_{n,k}\Gamma'^{-1}_k\otimes\Sigma_{e^{(h)},(m-n)}]=\Gamma_k^{-1}\Gamma_{(m-n),k}\Gamma'^{-1}_k\otimes\Sigma_{e^{(h)},(m-n)}$$

being constant

$$var(s_T) = (T-k-H)^{-1} \sum_{m=k}^{T-H} \sum_{|n-m|<h} E[(\Gamma_k^{-1} X_{m,k} \otimes I_r) vec((\varepsilon_{m+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{m+h-l}))$$

$$\times vec((\varepsilon_{n+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{n+h-l}))' (X'_{n,k} \Gamma_k'^{-1} \otimes I_r)]$$

$$= \sum_{|n-m|<h} E[(\Gamma_k^{-1} X_{m,k} \otimes I_r) vec((\varepsilon_{m+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{m+h-l})) vec((\varepsilon_{n+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{n+h-l}))' (X'_{n,k} \Gamma_k'^{-1} \otimes I_r)]$$

$$= \sum_{|n-m|<h} \Gamma_k^{-1} \Gamma_{(m-n),k} \Gamma_k'^{-1} \otimes \Sigma_{e^{(h)},(m-n)}$$

which is finite for finite h.

To show the Asymptotic negligibility of innovations, note that for $k > h-1$, the innovation is zero (this point ends up not mattering). Since $(\Gamma_k^{-1} X_{t,k} \otimes I_r) vec((\varepsilon_{t+h} + \sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}))$ is second order stationary (it has mean zero and it has been shown that the autocovariances are finite and constant at all horizons), then there exists a Wold VMA representation. This Wold representation can be written as a stationary VAR (∞). If I write the VAR (∞) as a VAR(1),

$$Z_t = AZ_{t-1} + e_t$$

so

$$r_{t0} = e_t$$

$$r_{t1} = Ae_t$$

$$r_{t2} = A^2 e_t$$

$$\vdots$$

Because the VAR is stationary, the impact of an innovation decays over time, and asymptotic negligibility trivially follows. □

A.4.5 Proof of Asymptotic Normality LP GLS

For LP OLS

$$\sqrt{T-k-H}[\hat{B}(k,h,OLS)-B(k,h)] = \sqrt{T-k-H}[U_{1T}\hat{\Gamma}_k^{-1} + U_{2T}\hat{\Gamma}_k^{-1} + U_{3T}\hat{\Gamma}_k^{-1}]$$

For LP GLS

$$\sqrt{T-k-H}[\hat{B}(k,h,GLS)-B(k,h)] = \sqrt{T-k-H}[U_{1T}\hat{\Gamma}_k^{-1}+U_{2T}\hat{\Gamma}_k^{-1}+U_{3T}\hat{\Gamma}_k^{-1}-U_{4T}\hat{\Gamma}_k^{-1}]$$

where again

$$U_{4T} = \{(T-k-H)^{-1}\sum_{t=k}^{T-H}(\sum_{l=1}^{h-1}\hat{\Theta}_l\hat{\varepsilon}_{t+h-l})X'_{t,k}\}$$

Theorem 3. *If assumption 3 holds, then for LP GLS*

$$\sqrt{T-k-H}vec[\hat{B}(k,h,GLS)-B(k,h)] \xrightarrow{d} N(0,\Omega_h^{GLS}),$$

where

$$\Omega_h^{GLS} = var(\aleph) + var(\Upsilon) + cov(\aleph,\Upsilon') + cov(\Upsilon,\aleph')$$

$$\aleph = (T-k-H)^{-1/2}vec[\sum_{t=k}^{T-H}\varepsilon_{t+h}X'_{t,k}\Gamma_k^{-1}]$$

$$\Upsilon = vec[\sum_{l=1}^{h-1}\Theta_l\sqrt{T-k-H}(\hat{B}(k,1)-B(k,1))\Gamma_{(h-l-1),k}\Gamma_k^{-1}]$$

Proof. To show that

$$\sqrt{T-k-H}[\hat{B}(k,h,GLS)-B(k,h)] = \sqrt{T-k-H}[U_{1T}\hat{\Gamma}_k^{-1}+U_{2T}\hat{\Gamma}_k^{-1}+U_{3T}\hat{\Gamma}_k^{-1}-U_{4T}\hat{\Gamma}_k^{-1}]$$

is normally distributed, it will first help to simplify the expression by showing

$$\sqrt{T-k-H}[\hat{B}(k,h,GLS)-B(k,h)] \xrightarrow{p} \sqrt{T-k-H}[U_{1T}\Gamma_k^{-1}+U_{2T}\Gamma_k^{-1}+U_{3T}\Gamma_k^{-1}-U_{4T}\Gamma_k^{-1}]$$

This can be done by showing that

$$\| \sqrt{T-k-H}[U_{1T}+U_{2T}+U_{3T}-U_{4T}](\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}) \| \xrightarrow{p} 0.$$

Jordà and Kozicki (2011) already showed that

$$\| \sqrt{T-k-H}[U_{1T}+U_{2T}+U_{3T}](\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}) \| \xrightarrow{p} 0.$$

So I just need to show

$$\| \sqrt{T-k-H}U_{4T}(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}) \| \xrightarrow{p} 0.$$

To simplify the expression into something more manageable, I'll begin by simplifying

$$\sqrt{T-k-H}[U_{4T}](\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}).$$

Let

$$\hat{\varepsilon}_t = \varepsilon_t + (\sum_{j=1}^{\infty} A_j y_{t-j}) - (\sum_{i=1}^{k} \hat{A}_i y_{t-i}),$$

then

$$\sqrt{T-k-H}U_{4T}(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}) = \{(T-k-H)^{-1/2}\sum_{t=k}^{T-H}(\sum_{l=1}^{h-1}\hat{\Theta}_l\hat{\varepsilon}_{t+h-l})X'_{t,k}\}(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}).$$

$$= \{(T-k-H)^{-1/2}\sum_{t=k}^{T-H}(\sum_{l=1}^{h-1}\hat{\Theta}_l(\varepsilon_{t+h-l}+(\sum_{j=1}^{\infty}A_jy_{t+h-l-j})-(\sum_{i=1}^{k}\hat{A}_iy_{t+h-l-i})))X'_{t,k}\}(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}).$$

$$= \{(T-k-H)^{-1/2}\sum_{t=k}^{T-H}(\sum_{l=1}^{h-1}\hat{\Theta}_l(\varepsilon_{t+h-l}+(\sum_{j=k+1}^{\infty}A_jy_{t+h-l-j})$$

$$-(\hat{B}(k,1)-B(k,1))X_{t+h-l-1,k}))X'_{t,k}\}(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}).$$

$$= \{(T-k-H)^{-1/2}\sum_{l=1}^{h-1}\hat{\Theta}_l\sum_{t=k}^{T-H}(\varepsilon_{t+h-l}+(\sum_{j=k+1}^{\infty}A_jy_{t+h-l-j})$$

$$-(\hat{B}(k,1)-B(k,1))X_{t+h-l-1,k}))X'_{t,k}\}(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}).$$

So

$$\| \sqrt{T-k-H}U_{4T}(\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}) \| \leq$$

$$\| \{(T-k-H)^{-1/2}\sum_{l=1}^{h-1}\hat{\Theta}_l\sum_{t=k}^{T-H}(\varepsilon_{t+h-l}+(\sum_{j=k+1}^{\infty}A_jy_{t+h-l-j})-(\hat{B}(k,1)-B(k,1))X_{t+h-l-1,k}))X'_{t,k}\} \|$$

$$\times \| (\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}) \|_1$$

$$\leq \sum_{l=1}^{h-1}\| \hat{\Theta}_l \| \Big(\| \{(T-k-H)^{-1/2}\sum_{t=k}^{T-H}(\varepsilon_{t+h-l}+(\sum_{j=k+1}^{\infty}A_jy_{t+h-l-j})$$

$$-(\hat{B}(k,1)-B(k,1))X_{t+h-l-1,k}))X'_{t,k}\} \| \Big)\{\| (\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}) \|_1\}$$

$$= \sum_{l=1}^{h-1}\| \hat{\Theta}_l \| \Big(\| \{[k(T-k-H)]^{-1/2}\sum_{t=k}^{T-H}(\varepsilon_{t+h-l}+(\sum_{j=k+1}^{\infty}A_jy_{t+h-l-j})$$

$$-(\hat{B}(k,1)-B(k,1))X_{t+h-l-1,k}))X'_{t,k}\} \| \Big)\{k^{1/2}\| (\hat{\Gamma}_k^{-1}-\Gamma_k^{-1}) \|_1\}$$

It has already been shown that

$$\| \hat{\Theta}_l \| \xrightarrow{p} \| \Theta_l \| < \infty,$$

for each $1 \leq l \leq h-1$. And we know from Lewis and Reinsel (1985) that $k^{1/2} \| (\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \|_1 \xrightarrow{p} 0$. Since $h-1$ is finite, to show

$$\| \sqrt{T-k-H} U_{4T} (\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \| \xrightarrow{p} 0,$$

I just need to show that

$$\Big(\| \{[k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\varepsilon_{t+h-l} + (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) - (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k})) X'_{t,k} \} \| \Big)$$

is bounded for each $1 \leq l \leq h-1$.

$$\Big(\| \{[k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\varepsilon_{t+h-l} + (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) - (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k})) X'_{t,k} \} \| \Big)$$

$$\leq \| [k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k} \| +$$

$$\| [k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\sum_{j=k+1}^{\infty} A_j y_{t+h-l-j}) X'_{t,k} \|$$

$$- \| [k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k} X'_{t,k} \| .$$

The first term is bounded since it was shown in the proof of consistency that

$$\| (T-k-H)^{-1} \sum_{t=k}^{T-H} \varepsilon_{t+h-l} X'_{t,k} \| = O_p(\Big(\frac{k}{T-k-H}\Big)^{1/2}).$$

Jordà and Kozicki (2011) show that the second term converges in probability to 0. For the

final term note that

$$\| \, [k(T-k-H)]^{-1/2} \sum_{t=k}^{T-H} (\hat{B}(k,1) - B(k,1)) X_{t+h-l-1,k} X'_{t,k} \, \|$$

$$\leq \underbrace{\left(\frac{T-k-H}{k}\right)^{1/2} \| \, (\hat{B}(k,1) - B(k,1)) \, \|}_{bounded} \underbrace{\| \, (T-k-H)^{-1} \sum_{t=k}^{T-H} X_{t+h-l-1,k} X'_{t,k} \, \|_1}_{bounded} .$$

Consequently

$$\| \, \sqrt{T-k-H} U_{4T} (\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \, \| \xrightarrow{p} 0,$$

and this completes the proof showing

$$\sqrt{T-k-H}[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p} \sqrt{T-k-H}[U_{1T}\Gamma_k^{-1} + U_{2T}\Gamma_k^{-1} + U_{3T}\Gamma_k^{-1} - U_{4T}\Gamma_k^{-1}].$$

From Jordà and Kozicki (2011) we know that

$$\| \, \sqrt{T-k-H} U_{1T} \Gamma_k^{-1} \, \| \xrightarrow{p} 0.$$

As a result

$$\sqrt{T-k-H}[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p} \sqrt{T-k-H}[U_{2T}\Gamma_k^{-1} + U_{3T}\Gamma_k^{-1} - U_{4T}\Gamma_k^{-1}].$$

Therefore

$$\sqrt{T-k-H}[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p} (T-k-H)^{-1/2} (\sum_{t=k}^{T-H} \varepsilon_{t+h} X'_{t,k}) \Gamma_k^{-1}$$

$$+ (T-k-H)^{-1/2} \sum_{t=k}^{T-H} (\sum_{l=1}^{h-1} \Theta_l \varepsilon_{t+h-l}) X'_{t,k} \Gamma_k^{-1}$$

$$-(T-k-H)^{-1/2}\sum_{t=k}^{T-H}(\sum_{l=1}^{h-1}\hat{\Theta}_l\hat{\varepsilon}_{t+h-l})X'_{t,k}\Gamma_k^{-1}.$$

Since

$$\hat{\varepsilon}_t = \varepsilon_t + (\sum_{j=1}^{\infty} A_j y_{t-j}) - (\sum_{i=1}^{k}\hat{A}_i y_{t-i})$$

$$= \varepsilon_t + (\sum_{j=k+1}^{\infty} A_j y_{t-j}) - (\hat{B}(k,1) - B(k,1))X_{t-1,k}),$$

it can be shown that

$$(T-k-H)^{-1/2}\sum_{t=k}^{T-H}(\sum_{l=1}^{h-1}\hat{\Theta}_l\hat{\varepsilon}_{t+h-l})X'_{t,k}\Gamma_k^{-1} \xrightarrow{p} (T-k-H)^{-1/2}\sum_{t=k}^{T-H}(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{t+h-l})X'_{t,k}\Gamma_k^{-1}$$

$$-\sum_{l=1}^{h-1}\Theta_l\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))\Gamma_{(h-l-1),k}\Gamma_k^{-1},$$

(the proof is omitted for brevity). Therefore

$$\sqrt{T-k-H}[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p} (T-k-H)^{-1/2}(\sum_{t=k}^{T-H}\varepsilon_{t+h}X'_{t,k})\Gamma_k^{-1}$$

$$+\sum_{l=1}^{h-1}\Theta_l\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))\Gamma_{(h-l-1),k}\Gamma_k^{-1}.$$

Note that

$$vec[(T-k-H)^{-1/2}\{\sum_{t=k}^{T-H}\varepsilon_{t+h}X'_{t,k}\}\Gamma_k^{-1} + \sum_{l=1}^{h-1}\Theta_l\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))\Gamma_{(h-l-1),k}\Gamma_k^{-1}]$$

$$= vec[(T-k-H)^{-1/2}\{\sum_{t=k}^{T-H}\varepsilon_{t+h}X'_{t,k}\}\Gamma_k^{-1}] + vec[\sum_{l=1}^{h-1}\Theta_l\sqrt{T-k-H}(\hat{B}(k,1) - B(k,1))\Gamma_{(h-l-1),k}\Gamma_k^{-1}]$$

$$= \{I_{kr}\otimes I_r\}vec[(T-k-H)^{-1/2}\{\sum_{t=k}^{T-H}\varepsilon_{t+h}X'_{t,k}\}\Gamma_k^{-1}]$$

$$+\Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)vec[\sqrt{T-k-H}(\hat{B}(k,1)-B(k,1))]$$

To show that

$$\sqrt{T-k-H}vec[\hat{B}(k,h,GLS)-B(k,h)]\xrightarrow{d}N(0,\Omega_h^{GLS}),$$

it suffices to show that the joint distribution of

$$\begin{bmatrix}\{I_{kr}\otimes I_r\}vec[(T-k-H)^{-1/2}\{\sum_{t=k}^{T-H}\varepsilon_{t+h}X'_{t,k}\}\Gamma_k^{-1}]\\ \Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)vec[\sqrt{T-k-H}(\hat{B}(k,1)-B(k,1))]\end{bmatrix}$$

converge to a normal distribution.

$$\begin{bmatrix}\{I_{kr}\otimes I_r\}vec[(T-k-H)^{-1/2}\{\sum_{t=k}^{T-H}\varepsilon_{t+h}X'_{t,k}\}\Gamma_k^{-1}]\\ \Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)vec[\sqrt{T-k-H}(\hat{B}(k,1)-B(k,1))]\end{bmatrix}$$

$$=\begin{bmatrix}\{I_{kr}\otimes I_r\} & 0\\ 0 & \Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)\end{bmatrix}\begin{bmatrix}vec[(T-k-H)^{-1/2}\{\sum_{t=k}^{T-H}\varepsilon_{t+h}X'_{t,k}\}\Gamma_k^{-1}]\\ vec[\sqrt{T-k-H}(\hat{B}(k,1)-B(k,1))]\end{bmatrix}$$

Define

$$s_T=\begin{bmatrix}\{I_{kr}\otimes I_r\} & 0\\ 0 & \Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)\end{bmatrix}\begin{bmatrix}vec[(T-k-H)^{-1/2}\{\sum_{t=k}^{T-H}\varepsilon_{t+h}X'_{t,k}\}\Gamma_k^{-1}]\\ vec[(T-k-H)^{-1/2}\{\sum_{t=k}^{T-H}\varepsilon_{t+1}X'_{t,k}\}\Gamma_k^{-1}]\end{bmatrix}.$$

To show asymptotic normality of s_T, the Gordin's CLT will be used. Using similar reasoning as for standard LP,

$$\{\begin{bmatrix}\{I_{kr}\otimes I_r\} & 0\\ 0 & \Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)\end{bmatrix}\begin{bmatrix}vec[\varepsilon_{t+h}X'_{t,k}\Gamma_k^{-1}]\\ vec[\varepsilon_{t+1}X'_{t,k}\Gamma_k^{-1}]\end{bmatrix}\}_{t=-\infty}^{t=\infty}$$

is a strongly stationary and ergodic sequence.

To show asymptotic normality for LP GLS, it needs to be shown that

$$s_T = \begin{bmatrix} \{I_{kr} \otimes I_r\} & 0 \\ 0 & \left(\sum_{l=1}^{h-1} \{\Gamma_k^{-1} \Gamma'_{(h-l-1),k} \otimes \Theta_l\} \right) \end{bmatrix} \begin{bmatrix} vec[(T-k-H)^{-1/2} \{\sum_{t=k}^{T-H} \varepsilon_{t+h} X'_{t,k}\} \Gamma_k^{-1}] \\ vec[(T-k-H)^{-1/2} \{\sum_{t=k}^{T-H} \varepsilon_{t+1} X'_{t,k}\} \Gamma_k^{-1}] \end{bmatrix}$$

is normally distributed. Since s_T is a strongly stationary and ergodic sequence, all that is left is to show the following conditions are satisfied:

1. Asymptotic uncorrelatedness

2. Summability of autocovariances

3. Asymptotic negligibility of innovations.

Asymptotic uncorrelatedness follows along the same lines as the standard LP and is omitted for brevity.

To show Summability of autocovariances, must show that

$$lim_{T \to \infty} var(s_T)$$

is finite and constant. Note that

$$\begin{bmatrix} \{I_{kr} \otimes I_r\} & 0 \\ 0 & \left(\sum_{l=1}^{h-1} \{\Gamma_k^{-1} \Gamma'_{(h-l-1),k} \otimes \Theta_l\} \right) \end{bmatrix} \begin{bmatrix} vec[\varepsilon_{t+h} X'_{t,k} \Gamma_k^{-1}] \\ vec[\varepsilon_{t+1} X'_{t,k} \Gamma_k^{-1}] \end{bmatrix}$$

$$= \begin{bmatrix} \{I_{kr} \otimes I_r\} & 0 \\ 0 & \left(\sum_{l=1}^{h-1} \{\Gamma_k^{-1} \Gamma'_{(h-l-1),k} \otimes \Theta_l\} \right) \end{bmatrix} \begin{bmatrix} (\Gamma_k^{-1} X_{t,k} \otimes I_r) vec[\varepsilon_{t+h}] \\ (\Gamma_k^{-1} X_{t,k} \otimes I_r) vec[\varepsilon_{t+1}] \end{bmatrix}$$

$$= \begin{bmatrix} \{I_{kr} \otimes I_r\} & 0 \\ 0 & \Big(\sum_{l=1}^{h-1} \{\Gamma_k^{-1} \Gamma'_{(h-l-1),k} \otimes \Theta_l\} \Big) \end{bmatrix} \begin{bmatrix} (\Gamma_k^{-1} X_{t,k} \otimes I_r)\varepsilon_{t+h} \\ (\Gamma_k^{-1} X_{t,k} \otimes I_r)\varepsilon_{t+1} \end{bmatrix}$$

$$= \begin{bmatrix} (\Gamma_k^{-1} X_{t,k} \otimes I_r)\varepsilon_{t+h} \\ \Big(\sum_{l=1}^{h-1} \{\Gamma_k^{-1} \Gamma'_{(h-l-1),k} \otimes \Theta_l\} \Big) (\Gamma_k^{-1} X_{t,k} \otimes I_r)\varepsilon_{t+1} \end{bmatrix}$$

The autocovariances for lag $m-n$ is

$$E[\begin{bmatrix} (\Gamma_k^{-1} X_{m,k} \otimes I_r)\varepsilon_{m+h} \\ l_k(\Gamma_k^{-1} X_{m,k} \otimes I_r)\varepsilon_{m+1} \end{bmatrix} \begin{bmatrix} (\Gamma_k^{-1} X_{n,k} \otimes I_r)\varepsilon_{n+h} \\ l_k(\Gamma_k^{-1} X_{n,k} \otimes I_r)\varepsilon_{n+1} \end{bmatrix}']$$

$$= E[\begin{bmatrix} (\Gamma_k^{-1} X_{m,k} \otimes I_r)\varepsilon_{m+h} \\ l_k(\Gamma_k^{-1} X_{m,k} \otimes I_r)\varepsilon_{m+1} \end{bmatrix} \begin{bmatrix} \varepsilon'_{n+h}(\Gamma_k^{-1} X_{n,k} \otimes I_r)' & \varepsilon'_{n+1}(\Gamma_k^{-1} X_{n,k} \otimes I_r)' l'_k \end{bmatrix}]$$

$$= E \begin{bmatrix} (\Gamma_k^{-1} X_{m,k} \otimes I_r)\varepsilon_{m+h}\varepsilon'_{n+h}(\Gamma_k^{-1} X_{n,k} \otimes I_r)' & (\Gamma_k^{-1} X_{m,k} \otimes I_r)\varepsilon_{m+h}\varepsilon'_{n+1}(\Gamma_k^{-1} X_{n,k} \otimes I_r)' l'_k \\ l_k(\Gamma_k^{-1} X_{m,k} \otimes I_r)\varepsilon_{m+1}\varepsilon'_{n+h}(\Gamma_k^{-1} X_{n,k} \otimes I_r)' & l_k(\Gamma_k^{-1} X_{m,k} \otimes I_r)\varepsilon_{m+1}\varepsilon'_{n+1}(\Gamma_k^{-1} X_{n,k} \otimes I_r)' l'_k \end{bmatrix}$$

where $l_k = \Big(\sum_{l=1}^{h-1} \{\Gamma_k^{-1} \Gamma'_{(h-l-1),k} \otimes \Theta_l\} \Big)$. Using the law of total expectations by first conditioning on the time n information set, $\mathcal{F}_n$, it follows that

$$E\Big(\begin{bmatrix} (\Gamma_k^{-1} X_{m,k} X'_{n,k} \Gamma_k^{-1} \otimes \Sigma_{\varepsilon,(m-n)}) & (\Gamma_k^{-1} X_{m,k} X'_{n,k} \Gamma_k^{-1} \otimes \Sigma_{\varepsilon,(m+1-n-h)}) l'_k \\ l_k(\Gamma_k^{-1} X_{m,k} X'_{n,k} \Gamma_k^{-1} \otimes \Sigma_{\varepsilon,(m+1-n-h)}) & l_k(\Gamma_k^{-1} X_{m,k} X'_{n,k} \Gamma_k^{-1} \otimes \Sigma_{\varepsilon,(m-n)}) l'_k \end{bmatrix} \Big)$$

$$= \Big(\begin{bmatrix} (\Gamma_k^{-1} \Gamma_{(m-n),k} \Gamma_k^{-1} \otimes \Sigma_{\varepsilon,(m-n)}) & (\Gamma_k^{-1} \Gamma_{(m-n),k} \Gamma_k^{-1} \otimes \Sigma_{\varepsilon,(m+1-n-h)}) l'_k \\ l_k(\Gamma_k^{-1} \Gamma_{(m-n),k} \Gamma_k^{-1} \otimes \Sigma_{\varepsilon,(m+1-n-h)}) & l_k(\Gamma_k^{-1} \Gamma_{(m-n),k} \Gamma_k^{-1} \otimes \Sigma_{\varepsilon,(m-n)}) l'_k \end{bmatrix} \Big)$$

which is finite since $\Gamma_k^{-1}, \Gamma_{(m-n),k}, \Sigma_{\varepsilon,(m-n)}, \Sigma_{\varepsilon,(m+1-n-h)}$, *and* l_k are bounded in probability. Therefore, the autocovariances of

$$\begin{bmatrix} (\Gamma_k^{-1} X_{t,k} \otimes I_r)\varepsilon_{t+h} \\ l_k(\Gamma_k^{-1} X_{t,k} \otimes I_r)\varepsilon_{t+1} \end{bmatrix}$$

are finite at all leads and lags.

For notational brevity let

$$q_m = \begin{bmatrix} (\Gamma_k^{-1} X_{m,k} \otimes I_r)\varepsilon_{m+h} \\ l_k(\Gamma_k^{-1} X_{m,k} \otimes I_r)\varepsilon_{m+1} \end{bmatrix}.$$

Note that

$$var(s_T) = (T-k-H)^{-1} \sum_{m=k}^{T-H} \sum_{n=k}^{T-H} E[q_m q_n'].$$

For $|n-m| \geq h$, $E[q_m q_n'] = 0$ due to independence so

$$var(s_T) = (T-k-H)^{-1} \sum_{m=k}^{T-H} \sum_{|n-m|<h} E[q_m q_n'].$$

Since the expectations are constant and $\frac{h}{T} \to 0$,

$$lim_{T\to\infty} var(s_T) = \sum_{|n-m|<h} E[q_m q_n'].$$

which is finite. This completes the proof of summability of autocovariances.

It was shown in the proof of summability of autocovariances that q_t is stationary. Hence asymptotic negligibility of innovations follows along the same lines as the proof of asymptotic normality for standard LP, so it is omitted for brevity. □

A.4.6 Proof of Asymptotic Efficiency LP GLS Relative to LP OLS

Theorem 4. *Under Assumption 3,*

$$var\{\sqrt{T-k-H}vec[\hat{B}(k,h,GLS)-B(k,h)]\}-var\{\sqrt{T-k-H}vec[\hat{B}(k,h,OLS)-B(k,h)]\}$$

is negative semi-definite. That is, the GLS estimator is more efficient than the OLS estimator.

Proof. The Wold representation can be inverted into an infinite order VAR representation

$$y_t = \sum_{j=1}^{\infty} A_j y_{t-j} + \varepsilon_t.$$

Any VAR(p) (including a VAR(∞)) can be written as a companion VAR(1). Denote this VAR(1) as

$$Y_t = AY_{t-1} + Z_t.$$

Take the eigenvalue decomposition of $A = E\Lambda E^{-1}$, where Λ is the diagonal matrix of distinct nonzero eigenvalues and E is the corresponding eigenmatrix and $EE^{-1} = I$ where I is the identity matrix. As a result $A^h = E\Lambda^h E^{-1}$. Define $W_t = E^{-1}Y_t$ and $\eta_t = E^{-1}Z_t$. This implies the VAR can be transformed into

$$W_t = \Lambda W_{t-1} + \eta_t.$$

Consequently

$$W_{t+h} = \Lambda^{h+1}W_{t-1} + \Lambda^h\eta_t + \ldots + \Lambda\eta_{t+h-1} + \eta_{t+h}.$$

Theorems 2 and 3 establish the consistency and asymptotic normality of LP OLS and LP GLS. If I can show the limiting distribution of GLS estimator is more efficient than the limiting distribution of OLS estimator for a stationary VAR(1) model at every horizon, it follows that the LP GLS estimator is asymptotically more efficient than the LP OLS estimator, since the mapping function from the LP estimates to the Wold coefficients is continuous and differentiable. Define

$$\sqrt{T-k-H}q = \sqrt{T-k-H}[\hat{B}(k,h,OLS) - \hat{B}(k,h,GLS)].$$

Note that

$$lim_{T\to\infty} var[\sqrt{T-H}vec\{\hat{B}(k,h,OLS) - B(k,h)\}]$$

$$= lim_{T\to\infty}\{var[\sqrt{T-H}vec\{\hat{B}(k,h,GLS) - B(k,h)\} + \sqrt{T-H}vec\{q\}]$$

$$= lim_{T\to\infty}\{var[\sqrt{T-k-H}vec[\hat{B}(k,h,GLS) - B(k,h)] + var[\sqrt{T-H}vec(q)]$$

$$+cov[\sqrt{T-k-H}[vec\{\hat{B}(k,h,GLS) - B(k,h)\}, \sqrt{T}vec\{q\}]$$

$$+cov[\sqrt{T-k-H}[vec\{\hat{B}(k,h,GLS) - B(k,h)\}, \sqrt{T}vec\{q\}]'\}.$$

To show that LP GLS is more efficient, it suffices to show that

$$lim_{T\to\infty} cov[\sqrt{T-k-H}vec\{q\}, \sqrt{T-k-H}vec\{\hat{B}(k,h,GLS) - B(k,h)\}] \geq 0,$$

in the positive semi-definite sense. Note that

$$\sqrt{T-k-H}[\hat{B}(k,h,GLS) - B(k,h)] \xrightarrow{p} (T-k-H)^{-1/2}(\sum_{t=k}^{T-H} \varepsilon_{t+h}X'_{t,k})\Gamma_k^{-1}$$

$$+\sum_{l=1}^{h-1}\Theta_l\sqrt{T-k-H}(\hat{B}(k,1)-B(k,1))\Gamma_{(h-l-1),k}\Gamma_k^{-1},$$

$$\sqrt{T-k-H}q \xrightarrow{p} \sqrt{T-k-H}U_{4T}\Gamma_k^{-1},$$

$$\sqrt{T-k-H}U_{4T}\Gamma_k^{-1} \xrightarrow{p} (T-k-H)^{-1/2}\sum_{t=k}^{T-H}(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{t+h-l})X'_{t,k}\Gamma_k^{-1}$$

$$-\sum_{l=1}^{h-1}\Theta_l\sqrt{T-k-H}(\hat{B}(k,1)-B(k,1))\Gamma_{(h-l-1),k}\Gamma_k^{-1}.$$

So

$$lim_{T\to\infty}cov[\sqrt{T-k-H}vec\{U_{4T}\Gamma_k^{-1}\},\sqrt{T-k-H}vec\{\hat{B}(k,h,GLS)-B(k,h)\}]$$

$$=(T-k-H)^{-1}\sum_{m=k}^{T-H}\sum_{n=k}^{T-H}E[(\Gamma_k^{-1}X_{m,k}\otimes I_r)(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})\varepsilon'_{n+h}(\Gamma_k^{-1}X_{n,k}\otimes I_r)']$$

$$+(T-k-H)^{-1}\sum_{m=k}^{T-H}\sum_{n=k}^{T-H}E[(\Gamma_k^{-1}X_{m,k}\otimes I_r)(\sum_{l=1}^{h-1}\Theta_l\varepsilon_{m+h-l})$$

$$\times\varepsilon'_{n+1}(\Gamma_k^{-1}X_{n,k}\otimes I_r)'\Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)']$$

$$-(T-k-H)^{-1}\sum_{m=k}^{T-H}\sum_{n=k}^{T-H}E[\Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)(\Gamma_k^{-1}X_{m,k}\otimes I_r)\varepsilon_{m+1}\varepsilon'_{n+h}(\Gamma_k^{-1}X_{n,k}\otimes I_r)']$$

$$-(T-k-H)^{-1}\sum_{m=k}^{T-H}\sum_{n=k}^{T-H}E[\Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)(\Gamma_k^{-1}X_{m,k}\otimes I_r)\varepsilon_{m+1}$$

$$\times\varepsilon'_{n+1}(\Gamma_k^{-1}X_{n,k}\otimes I_r)'\Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)'].$$

By independence and since the expectations are finite

$$=\sum_{l=1}^{h-1}\Big((\Gamma_k^{-1}\Gamma'_{k,(-l)}\Gamma_k^{-1}\otimes\Theta_l\Sigma_\varepsilon)\Big)$$

$$+\sum_{l=1}^{h-1}\Big((\Gamma_k^{-1}\Gamma'_{k,(h-l-1)}\Gamma_k^{-1}\otimes\Theta_l\Sigma_\varepsilon)\Big)\Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)'$$

$$-[\Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)(\Gamma_k^{-1}\Gamma'_{(h-1),k}\Gamma_k^{-1}\otimes\Sigma_\varepsilon)]$$

$$-[\Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)(\Gamma_k^{-1}\otimes\Sigma_\varepsilon)\Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)'].$$

Second and fourth lines cancel so

$$lim_{T\to\infty}cov[\sqrt{T-k-H}vec\{U_{4T}\Gamma_k^{-1}\},\sqrt{T-k-H}vec\{\hat{B}(k,h,GLS)-B(k,h)\}]$$

$$=\sum_{l=1}^{h-1}\Big((\Gamma_k^{-1}\Gamma'_{k,(-l)}\Gamma_k^{-1}\otimes\Theta_l\Sigma_\varepsilon)\Big)$$

$$-[\Big(\sum_{l=1}^{h-1}\{\Gamma_k^{-1}\Gamma'_{(h-l-1),k}\otimes\Theta_l\}\Big)(\Gamma_k^{-1}\Gamma'_{(h-1),k}\Gamma_k^{-1}\otimes\Sigma_\varepsilon)].$$

In the case where the true model can be written as a VAR(1) which has been diagonalized then

$$lim_{T\to\infty}cov[\sqrt{T-k-H}vec\{U_{4T}\Gamma_k^{-1}\},\sqrt{T-k-H}vec\{\hat{B}(k,h,GLS)-B(k,h)\}]$$

$$=\sum_{l=1}^{h-1}\Big((\Gamma_w^{-1}\Gamma_w\Lambda'^l\Gamma^{-1}\otimes\Lambda^l\Sigma_\eta)\Big)-[\Big(\sum_{l=1}^{h-1}\{\Gamma_w^{-1}\Gamma'_w\Lambda'^{h-l-1}\otimes\Lambda^l\}\Big)(\Gamma_w^{-1}\Gamma'_w\Lambda'^{h-1}\Gamma_w^{-1}\otimes\Sigma_\eta)]$$

$$=\sum_{l=1}^{h-1}\Big((\Lambda'^l\Gamma_w^{-1}\otimes\Lambda^l\Sigma_\eta)\Big)-[\Big(\sum_{l=1}^{h-1}\{\Lambda'^{h-l-1}\otimes\Lambda^l\}\Big)(\Lambda'^{h-1}\Gamma_w^{-1}\otimes\Sigma_\eta)].$$

$$=\sum_{l=1}^{h-1}\Big((\Lambda'^l\Gamma_w^{-1}\otimes\Lambda^l\Sigma_\eta)\Big)-[\Big(\sum_{l=1}^{h-1}\{\Lambda^{2h-l-2}\Gamma_w^{-1}\otimes\Lambda^l\Sigma_\eta\}\Big)]$$

$$=\sum_{l=1}^{h-1}\Big((\Lambda'^l-\Lambda^{2h-l-2})\Gamma_w^{-1}\otimes\Lambda^l\Sigma_\eta)\Big)$$

where $E(W_tW_t')=\Gamma_w$, and since the model is a VAR(1), $E(W_tW_{t-j}')=\Lambda^j\Gamma_w$. Note that the dimensions of the parameters have been suppressed for simplicity. Premultiply corresponding

terms in the sum by identity matrix

$$(\Lambda^{l} \otimes \Lambda^{-l}) = I$$

yields

$$\sum_{l=1}^{h-1} \Big((\Lambda^{2l} - \Lambda^{2h-2}) \Gamma_w^{-1} \otimes \Sigma_\eta \Big)$$

$$= \sum_{l=1}^{h-1} (\Lambda^{2l} \Gamma_w^{-1} \otimes \Sigma_\eta) - (h-1)(\Lambda^{2(h-1)} \Gamma_w^{-1} \otimes \Sigma_\eta)$$

which is positive definite since

$$\sum_{l=1}^{h-1} (\Lambda^{2l-2(h-1)} \Lambda^{2(h-1)} \Gamma_w^{-1} \otimes \Sigma_\eta) - (h-1)(\Lambda^{2(h-1)} \Gamma_w^{-1} \otimes \Sigma_\eta) \geq 0$$

$$\sum_{l=1}^{h-1} (\Lambda^{-2(h-l-1)} \Lambda^{2(h-1)} \Gamma_w^{-1} \otimes \Sigma_\eta) - (h-1)(\Lambda^{2(h-1)} \Gamma_w^{-1} \otimes \Sigma_\eta) \geq 0$$

$$\sum_{l=1}^{h-1} (\Lambda^{-2(h-l-1)} \Lambda^{2(h-1)} \Gamma_w^{-1} \otimes \Sigma_\eta) [(h-1)(\Lambda^{2(h-1)} \Gamma_w^{-1} \otimes \Sigma_\eta)]^{-1} - I \geq 0$$

$$(h-1)^{-1} \sum_{l=1}^{h-1} (\Lambda^{-2(h-l-1)} \otimes I) - I \geq 0 \; for \; h = 2, 3, \ldots$$

since Λ is diagonal matrix where all of diagonal elements are less than one is absolute value and $l + 1 \leq h$. Therefore GLS is more efficient since

$$lim_{T \to \infty} var[\sqrt{T-k-H}[\hat{B}(k,h,OLS) - B(k,h)]$$

$$= lim_{T \to \infty} \{ \underbrace{var[\sqrt{T-k-H}[\hat{B}(k,h,GLS) - B(k,h)]}_{pos} + \underbrace{var[\sqrt{T-H}q]}_{pos}$$

$$+ \underbrace{cov[\sqrt{T-k-H}[\hat{B}(k,h,GLS) - B(k,h), \sqrt{T}q]}_{pos-semi}$$

$$+\underbrace{cov[\sqrt{T-k-H}[\hat{B}(k,h,GLS)-B(k,h),\sqrt{T}q]'}_{pos-semi}\}.$$

$\square$

A.5 Figures

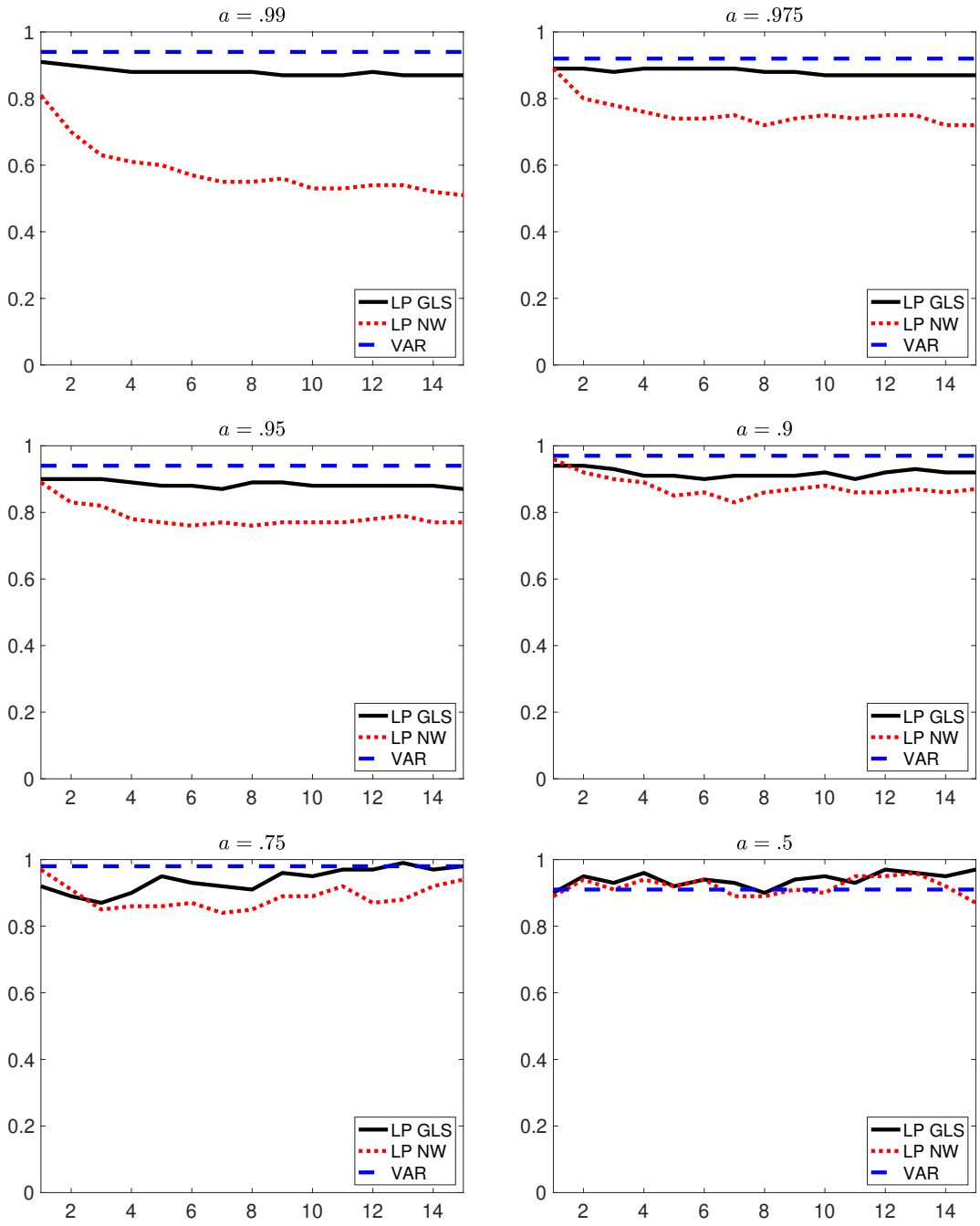

Figure A.1: Coverage Rates for 95% Confidence Intervals

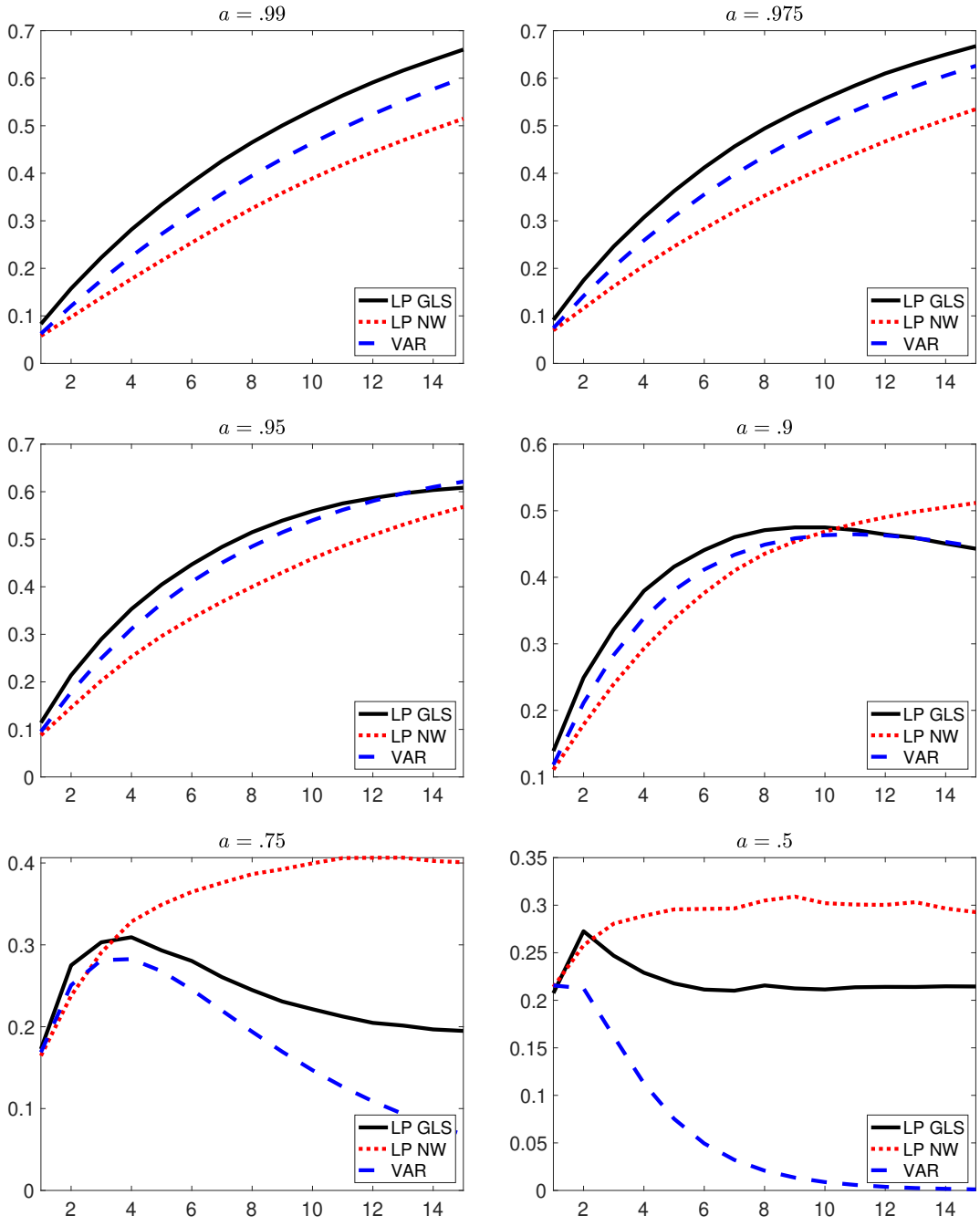

Figure A.2: Average Length for 95% Confidence Intervals

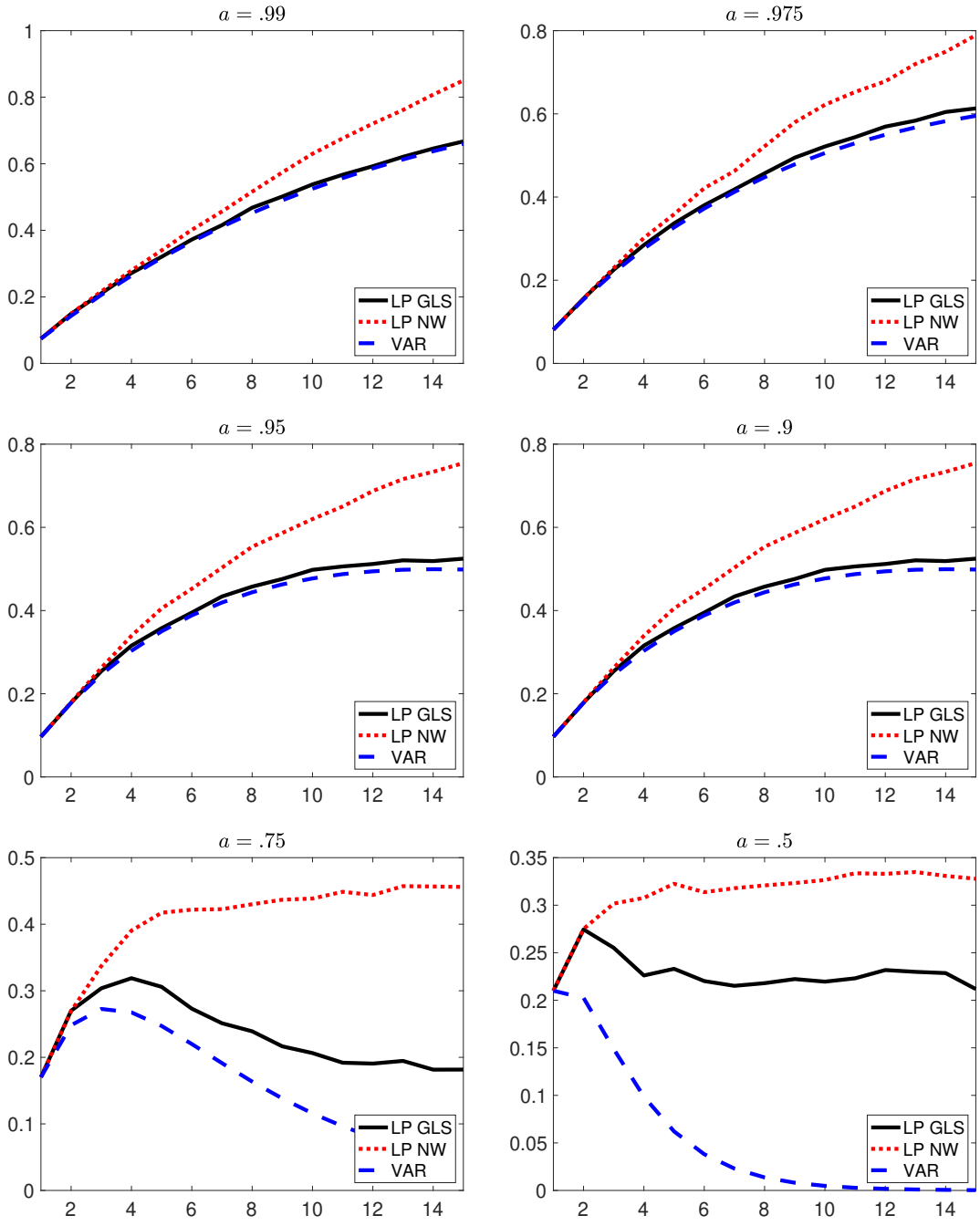

Figure A.3: Monte Carlo Simulation of "True" Length for 95% Confidence Intervals

Appendix B

Appendix to Chapter 2

B.1 Data

All of the data was downloaded from the Ramey (2016) database, which can be found here: http://econweb.ucsd.edu/~vramey/research.html#data.

GDP and GDP Deflator:

Quarterly data from 1890-2015

Government Expenditure:

Quarterly data from 1890-2015

Defense News:

Defense news variable from 1890-2015.

Population:

Tax Revenues:

Quarterly data from 1890-2015 on nominal "Government Current Receipts"

Total Deficit:

1890-2015

Interest Rate:

1890-2015

B.2 Time-Varying Parameter Model

Time-varying parameter models are modeled as in Prado and West (2010). Let

$$y_t = X_t'\beta_t + v_t,$$

$$\beta_t = \beta_{t-1} + w_t,$$

where y_t is a $m \times 1$ vector β_t is the $p \times 1$ state vector at time t, X_t is a $p \times m$ vector of regressors at time t, v_t is a $m \times 1$ vector observation noise with $v_t \sim N(0, \Sigma_t)$, w_t is the

state evolution noise with $w_t \sim N(0, \Sigma_t \otimes W_t)$, and v_s and w_t are independent and mutually independent $\forall s, t$. Notice that the variance of v_t is allowed to be time-varying. Stochastic volatility (time-varying variance) is modeled as a beta-Bartlett Wishart random walk. Define D_{t-1} is the amount of information known at time $t-1$. The beta-Bartlett Wishart random walk is defined using the following $t-1$ to time t update

$$p(\Sigma_{t-1}|D_{t-1}) \sim IW(n_{t-1}, \Psi_{t-1})$$

and

$$p(\Sigma_t|D_{t-1}) \sim IW(\theta n_{t-1}, b_t \Psi_{t-1}),$$

where θ is a discount factor for stochastic volatility and $b_t = (\theta n_{t-1} + k - 1)/(n_{t-1} + k - 1)$.[1] The models are estimated using discount factors and the Forward Filter Backward Sampler (FFBS) algorithm, and details about the estimation procedure can be found in the Appendix.[2] Because discount factors and conjugate priors are used, MCMC is not needed. This is crucial for three reasons. First, if the number or parameters is even moderately large, time-varying parameter models such as Cogley and Sargent (2005); Primiceri (2005) become computationally demanding to estimate if not infeasible (Koop and Korobilis, 2013). Second, LP are estimated horizon by horizon in a sequential fashion which can make procedures such as Cogley and Sargent (2005); Primiceri (2005) impractical. Third, in order to do model comparison or hypothesis testing, it is often necessary to calculate the marginal likelihood, which is no trivial task for models estimated using MCMC. In recent years discount factors have been used in the as a solution to when the procedures of Cogley and Sargent (2005); Primiceri (2005) are burdensome (Koop and Korobilis, 2013; Koop et al., 2018). This is not to suggest that time-varying parameter procedures such as Cogley and Sargent (2005);

[1] The model uses different discount factors for the regression coefficients and stochastic volatility.

[2] See West and Harrison (1997); Prado and West (2010) for derivations and more details about time-varying parameter methods using discount factors.

Primiceri (2005) or other cannot be used, just that depending on the goal of the analysis and the computational power available to the researcher, these procedures may not be practical.[3]

Discount factors (also known as forgetting factors) are a natural framework for allowing and controlling for time variation in regression coefficients and the variance and are a core part of the Bayesian forecasting literature (West and Harrison, 1997; Prado and West, 2010). Discount factors lie in the interval $(0, 1]$. If a discount factor, say $\theta = .99$ is used, then from period $t \to t+1$, $\frac{1}{\theta} - 1 \approx 1\%$ of information known at time t is discounted or forgotten in the Kalman filtration process.[4] And if $\theta = .99$, observations from 20 periods ago receive approximately 80% as much weight as this period's observation. The loss of information over time allows more recent data to have a larger impact on the parameter value and is the crux for controlling for time variation in the parameters. The discount factors are estimated using Griddy Gibbs. Including the the discount factor as a parameter to be estimated takes into account uncertainty in the hyperparameters and is a natural way to safeguard against overfitting (Giannone et al., 2015).

Due to the number of parameters being estimated, the priors for time-varying parameter models are quite important (Koop and Korobilis, 2009), otherwise parameter estimates may be imprecise if the sample size is not large. Like Cogley and Sargent (2005); Primiceri (2005), a training sample prior can be used.

The conjugate training sample prior for $p(\beta_1, \Sigma_1)$ is

$$p(\beta_1|\Sigma_1) \sim N(\underline{b}, \Sigma_u^{(0)} \otimes \underline{\Omega}),$$

[3] If time-varying parameter procedures such as Cogley and Sargent (2005); Primiceri (2005) are used, it is recommended that the MCMC be implemented using the more computationally efficient precision sampler in Chan and Jeliazkov (2009).

[4] A discount factor of .99 has properties similar to what Cogley and Sargent (2005) call their "business as usual" prior, and it can be shown that the choice of prior shrinkage coefficient in Cogley and Sargent (2005) allows for variation in the regression coefficients roughly similar to that allowed for by a regression coefficient discount factor of .99 (Koop and Korobilis, 2013).

$$p(\Sigma_1) \sim IW(\underline{n}, \underline{\Psi}).$$

$\underline{n}$ is the prior degrees of freedom, $\underline{b} = \hat{\beta}_{OLS}$ and $\underline{\Psi} = \underline{n}\hat{\Sigma}_{OLS}$, where $\hat{\beta}_{OLS}$ and $\hat{\Sigma}_{OLS}$ are the OLS results from the training sample. $\underline{\Omega} = \frac{T}{\underline{n}}(\underline{X}'\underline{X})^{-1}$ where $\underline{X}$ is the design matrix for the training sample and $\frac{T}{\underline{n}}$ rescales the conditional variance of β_1 so the conditional distribution will have the asymptotic variance of the OLS results based on the average of $\underline{n}$ observations.[5] $\underline{n}$, which determines the informativeness of the prior, can be chosen by the researcher or a prior can be placed on $\underline{n}$ and estimated using Griddy Gibbs or sampling importance resampling. In order for the prior mean of Σ_0 to be defined, $\underline{n} \geq p + 2$. By default, I set $\underline{n} = p + 2$ to make the prior weakly informative but still proper. The diagonal of $\underline{\Omega}$ can be taken to prevent collinearity issues if the prior is only based on small training sample (Brodersen et al., 2015).

B.3 Forward Filter Backward Sampler (FFBS)

B.3.1 Forward Filtering

More detail about the algorithm can be found in Prado and West (2010). Recall that a TVP model can be characterized as follows:

$$y_t = X_t'\beta_t + v_t,$$

$$\beta_t = \beta_{t-1} + w_t,$$

where y_t is a $m \times 1$ vector β_t is the $p \times 1$ state vector at time t, X_t is a $p \times m$ vector of regressors at time t, ϵ_t is a $m \times 1$ vector observation noise with $v_t \sim N(0, \Sigma_t)$, w_t is the state evolution

[5] This is in the spirit of the unit information prior (Kass and Wasserman, 1995), but since this is done over a training sample, it does not make double use of the data.

noise with $w_t \sim N(0, \Sigma_t \otimes W_t)$, and v_s and w_t are independent and mutually independent $\forall s, t$. Notice that the variance of v_t is allowed to be time-varying. Stochastic volatility (time-varying variance) is modeled as a beta-Bartlett Wishart random walk. Stochastic volatility is modeled as a beta-Bartlett Wishart random walk which is defined as following $t-1$ to time t update

$$p(\Sigma_{t-1}|D_{t-1}) \sim IW(n_{t-1}, \Psi_{t-1})$$

then

$$p(\Sigma_t|D_{t-1}) \sim IW(\theta n_{t-1}, b_t \Psi_{t-1})$$

where θ is a discount factor for stochastic volatility and $b_t = (\theta n_{t-1}+k-1)/(n_{t-1}+k-1)$. Let D_0 represents initial prior information and the current information set represented by $D_t = \{D_{t-1}, y_t\}$. The estimates of a standard TVP DLM can be obtained as follows. First recall that for a VAR(k) $X_t \equiv I_n \otimes [y'_{t-1}, \ldots, y'_{t-k}]$. Imagine we have the posterior distributions of β_t and v_t at time $t-1$. The posteriors are:

$$\beta_{t-1}|\Sigma_{t-1}, D_{t-1} \sim N(m_{t-1}, \Sigma_{t-1} \otimes C_{t-1}),$$

$$\Sigma_{t-1}|D_{t-1} \sim IW(n_{t-1}, \Psi_{t-1}),$$

where

$$M_t = M_{t-1} + A_t \epsilon'_t,$$

$$m_t = vec(M_t),$$

$$C_t = R_t - A_t A'_t q_t,$$

$$A_t = R_t X_t / q_t,$$

$$R_t = C_{t-1} + W_t = C_{t-1}/\delta.$$

$$n_t = \theta n_{t-1} + 1,$$

$$\Psi_t = \Psi_{t-1} + \epsilon_t \epsilon_t^{'} / q_t,$$

$$\epsilon_t = y_t - f_t,$$

$$f_t = X_t' M_{t-1},$$

$$q_t = X_t' R_t X_t' + 1,$$

where δ is the discount factor for the regression coefficients.The volatility evolves from the Σ_{t-1} posterior to the prior of Σ_t according to

$$p(\Sigma_t | D_{t-1}) \sim IW(\theta n_{t-1}, \theta \Psi_{t-1})$$

State evolves from the β_{t-1} prior to the β_t posterior as follows:

$$\beta_t | \Sigma_t, D_{t-1} \sim N(m_{t-1}, \Sigma_t \otimes R_t),$$

$\beta_t | D_{t-1}$ and $\Sigma_t | D_{t-1}$ are now the priors for β_t and Σ_t respectively. This leads to the following one-step-ahead predictive of y_t:

$$y_t | D_{t-1} \sim T_{\theta n_{t-1}}(f_t, q_t \frac{\Psi_{t-1}}{n_{t-1}}),$$

where The posterior for $\beta_t | D_t$ and $\Sigma_t | D_t$ can be now be calculated.

B.3.2 Backward Sampling

Initialize at T draw

$$\Sigma_T | D_T \sim IW(n_t, \Psi_t),$$

$$\beta_T | \Sigma_T, D_T \sim N(M_T, \Sigma_T \otimes C_T).$$

For $t-1$ to 1

$$\Sigma_t^{-1} = \theta \Sigma_{t+1}^{-1} + \gamma_t,$$

where

$$\gamma_t^{-1} \sim IW((1-\theta)n_t, \Psi_t),$$

and

$$\beta_t = m_t + \delta(\beta_{t+1} - m_t) + N(0, \Sigma_t \otimes C_t^*),$$

where

$$C_t^* = C_t - \delta^2 R_{t+1}.$$

B.4 Choosing Lag Length and Estimating Discount Factors

The optimal the lag length is chosen by maximizing the joint log likelihood functions defined in terms of the predictive densities

$$\log[p(y_{1:T}|D_0, \delta, \theta, lag\, length)] = \sum_{t=1}^{T} \log[p(y_t|D_{t-1}, \delta, \theta, lag\, length)],$$

where

$$p(y_t|D_{t-1}, \delta, \theta, lag\, length),$$

is the one step ahead predictive density, δ is the discount factor that controls for time variation in the regression coefficients, and D_{t-1} is the amount of information known at time $t-1$.[6] Maximizing the joint log likelihood functions is equivalent to maximizing the marginal likelihood. If each model is assumed to have the same prior probability, it is also equivalent to choosing the model with the highest posterior probability. Let $M_1, M_2, \ldots, M_I$ denote I models of the same structure that only differ in their lag lengths. The posterior probability for model i can be calculated by:

$$p(M_i|y_{1:T}, D_0) = \frac{p(M_i)p(y_{1:T}|D_0, M_i)}{\sum_{j=1}^{I} p(M_j)p(y_{1:T}|D_0, M_j)}.$$

Assuming all models have equal prior probability ($p(M_i) = I^{-1}\, \forall i$):

$$p(M_i|y_{1:T}, D_0) = \frac{p(y_{1:T}|D_0, M_i)}{\sum_{j=1}^{I} p(y_{1:T}|D_0, M_j)}.$$

[6] X_t is suppressed in the marginal likelihood for clarity.

Then conditional on the optimal lag length, the posterior distributions for the regression coefficients and the variance are model averaged over the grid of discount factors in order to take into account the uncertainty in the discount factors. Model averaging over the grid of discount factors is equivalent to placing a uniform prior on the discount factors and estimating them using Griddy Gibbs. Ideally one would use sampling importance resampling (see (Lopes et al., 1999)), but this is computationally impractical.

The regression coefficients' discount factor is estimated over a default grid of [.7, 1] where the grid is partitioned by .01. The stochastic volatility discount factor is also chosen over a default grid of [.7, 1] where the grid is partitioned by .01. The initial grid size and partition are chosen because they cover fairly rapid parameter changes to no parameter change and should cover most situations (West and Harrison, 1997) .[7] It is important to note that if posterior distribution of the discount factors pile up at the bottom of the grid, the grid must be lowered. For example let us say that the median regression coefficient discount factor is .95, but the median variance discount factor is .7. The grid for the variance discount factor must be lowered (e.g., to [.6, 1]). The reason for this is because the true discount factor for the variance may be .62 and the regression coefficient discount factor 1, but because the grid initially only searched over [.7, 1], it may be optimal for the regression coefficients to allow for time variation in order to compensate for the bound on the amount of stochastic volatility. Theoretically, one could allow just for regression coefficient instability or only for stochastic volatility. One would just have to restrict the discount factor not of interest to be equal to 1 and then search the grid for the other discount factor. This is not recommended because the restriction may exaggerate the results of the test. For example, if the true model has stochastic volatility and the test is restricted not to allow for stochastic volatility, it may be optimal for the time-varying parameter model to exaggerate the amount of time variation in the regression coefficients in order to compensate for the restriction.[8]

[7] Depending on the context, these grid values may not be appropriate and can be adjusted accordingly. If desired, one can also conduct a sensitivity analysis with the size of the grid partitions.

[8] A similar argument is made by Sims and Zha (2006) on an earlier version of Cogley and Sargent (2005)

B.5 Figures

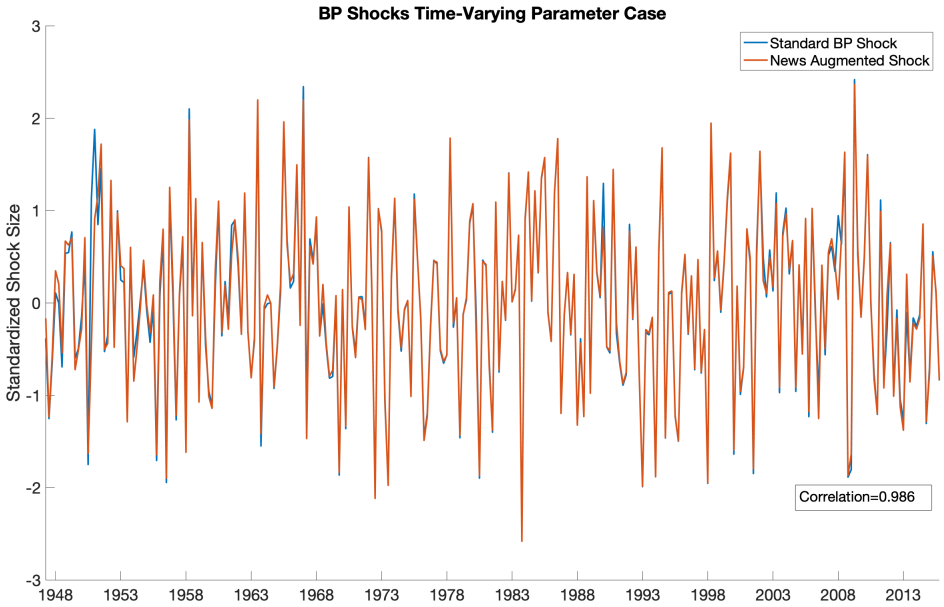

Figure B.1: Blanchard-Perotti Shocks for Time-Varying Parameter Models (Ramey Case)

that did not allow for stochastic volatility in their time-varying parameter model.

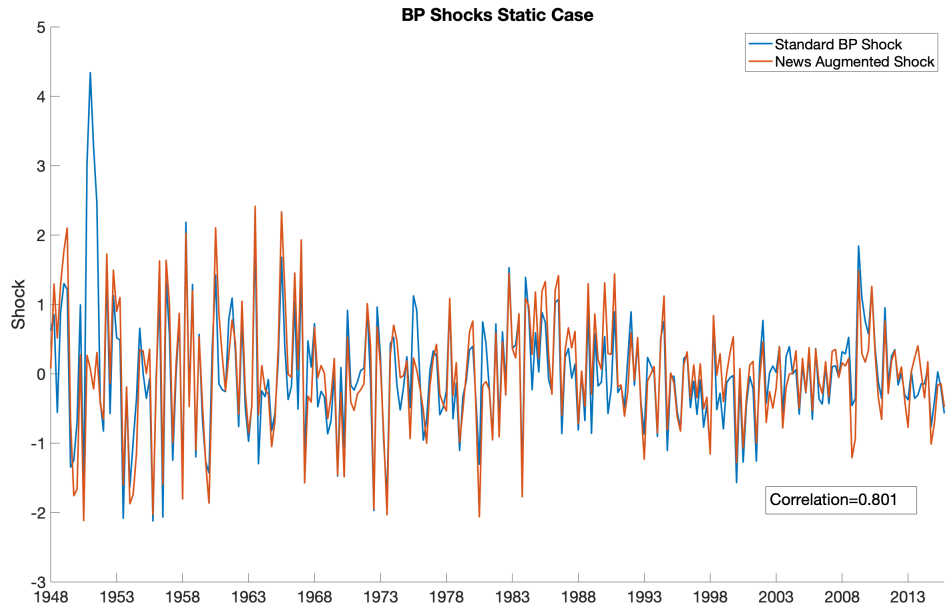

Figure B.2: Blanchard-Perotti Shocks for Static Models (Ramey Case)

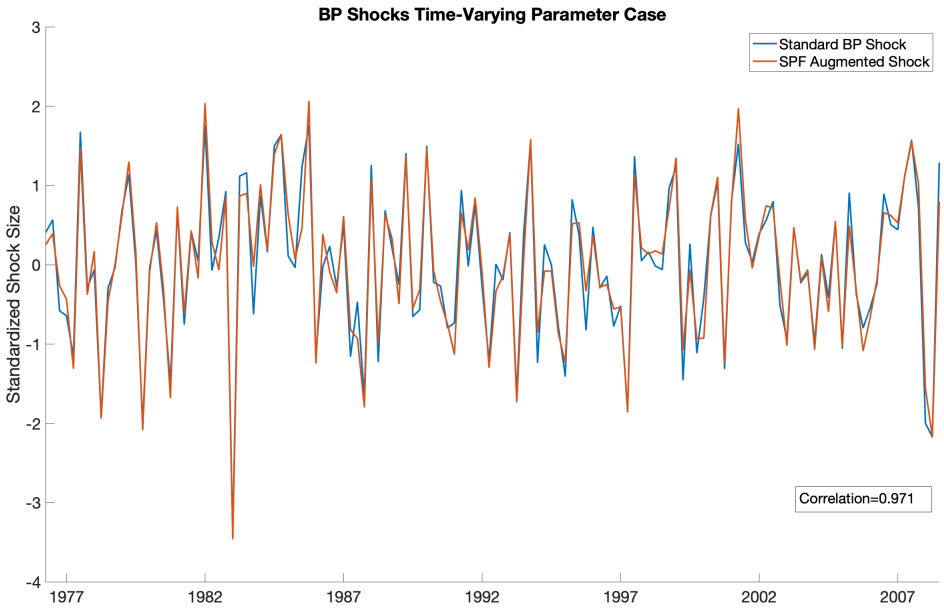

Figure B.3: Blanchard-Perotti Shocks for Time-Varying Parameter Models (AG Case)

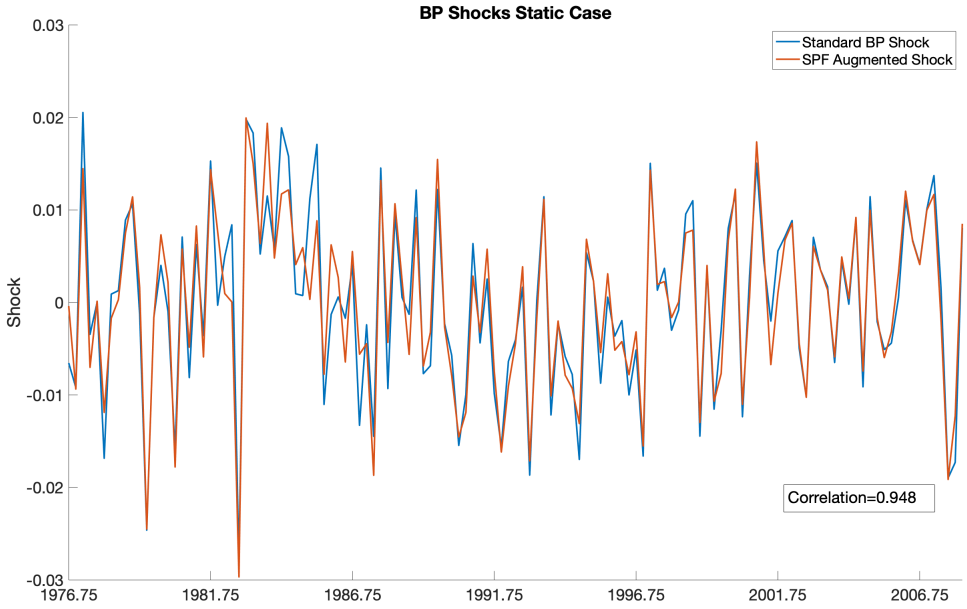

Figure B.4: Blanchard-Perotti Shocks for Time-Varying Parameter Models (AG Case)

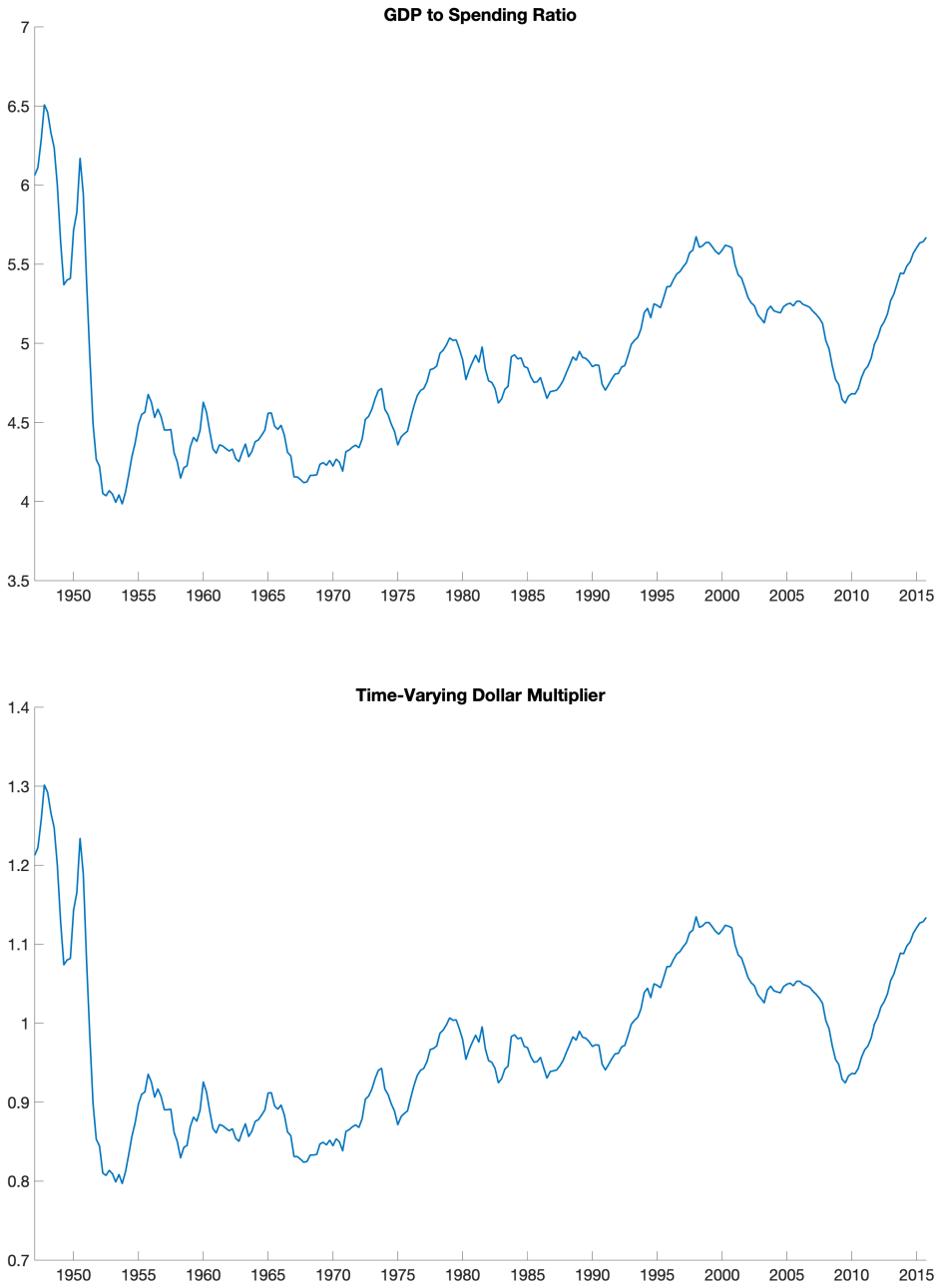

Figure B.5: GDP to Spending Ratio and Hypothetical Dollar Multiplier Plot

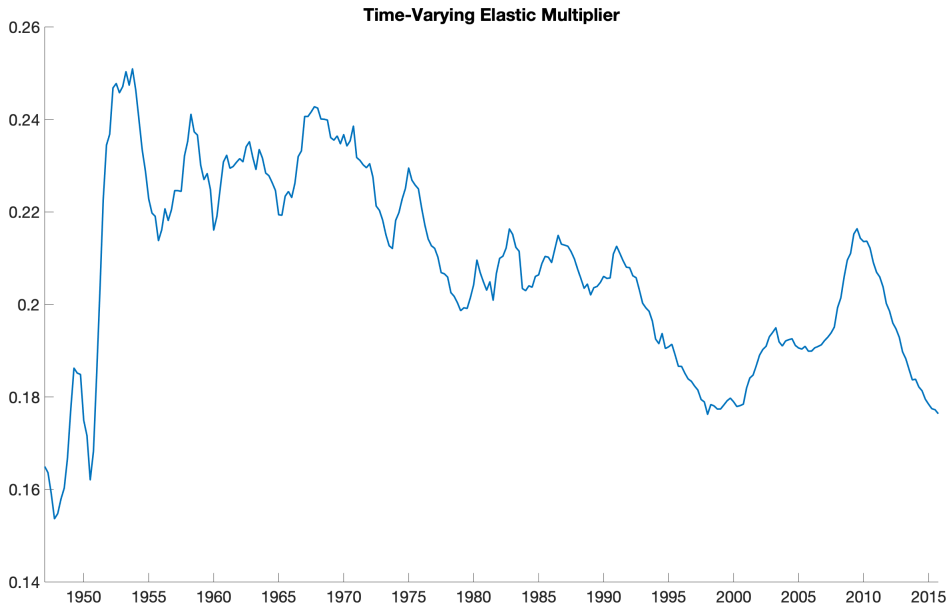

Figure B.6: Hypothetical Elastic Multiplier Plot

Appendix C

Appendix to Chapter 3

C.1 Data

Unless stated otherwise, the data was downloaded from the St. Louis Federal Reserve Database.

GDP :

Quarterly data from 1947Q1-2007Q4 on real GDP.

Interest Rate:

Quarterly data from 1947Q1-2007Q4 3-month Treasury bill.

PCE Deflator:

Quarterly data from 1947Q1-2007Q4 Personal Consumption Expenditures.

Smets and Wouters (2007) Data:

The data is downloaded from: https://www.aeaweb.org/articles?id=10.1257/aer.97.3.586.

C.2 Forward Filtering Algorithm

C.2.1 Forward Filtering

More detail about the algorithm can be found in West and Harrison (1997); Prado and West (2010). Recall that a TVP model can be characterized as follows:

$$y_t = X_t'\beta_t + \varepsilon_t,$$

$$\beta_t = \beta_{t-1} + w_t,$$

where β_t is the $p \times 1$ state vector at time t. X_t is a p-dimensional vector of regressors at time t. ε_t is the observation noise at time t with $\varepsilon_t \sim N(0, v_t)$, and w_t is the state evolution noise with $w_t \sim T_{n_{t-1}}(0, W_t)$. Stochastic volatility is modeled as a Beta/Gamma random walk.

$$v_t = v_{t-1}/(\theta/\gamma_t),$$

where

$$\gamma_t \sim Beta(\frac{\theta n_{t-1}}{2}, \frac{(1-\theta)n_{t-1}}{2}),$$

and θ is a discount factor for stochastic volatility. Let D_0 represents initial prior information and the current information set represented by $D_t = \{D_{t-1}, y_t\}$. The estimates of a standard TVP DLM can be obtained as follows. Imagine we have the posterior distributions of β_t and v_t at time $t-1$. The posteriors are:

$$\beta_{t-1}|v_{t-1}, D_{t-1} \sim N(m_{t-1}, \frac{v_{t-1}}{s_{t-1}} C_{t-1}),$$

$$v_{t-1}|D_{t-1} \sim n_{t-1} s_{t-1} / \chi^2_{n_{t-1}},$$

where

$$m_t = m_{t-1} + A_t e_t,$$

$$s_t = r_t s_{t-1},$$

$$r_t = (\theta n_{t-1} + e_t^2 / q_t) / n_t,$$

$$n_t = \theta n_{t-1} + 1,$$

$$C_t = r_t (R_t - A_t A_t' q_t).$$

The volatility evolves from the v_{t-1} posterior to the prior of v_t according to

$$v_t|D_{t-1} \sim (\theta n_{t-1}) s_{t-1} / \chi^2_{\theta n_{t-1}}.$$

State evolves from the β_{t-1} prior to the β_t posterior as follows:

$$\beta_t|v_t, D_{t-1} \sim N(m_{t-1}, \frac{v_t}{s_{t-1}} R_t),$$

where

$$R_t = C_{t-1} + W_t = C_{t-1}/\delta.$$

δ is the discount factor for the regression coefficients. $\beta_t|v_t, D_{t-1}$ and $v_t|D_{t-1}$ are now the priors for β_t and v_t respectively. Marginalizing out v_t from $\beta_t|v_t, D_{t-1}$ leads to the following marginal prior:

$$\beta_t|D_{t-1} \sim T_{\theta n_{t-1}}(m_{t-1}, R_t),$$

where $T_{\theta n_{t-1}}(m_{t-1}, R_t)$ is a T distribution with θn_{t-1} degrees of freedom, location a_t and scale R_t. This leads to the following one-step-ahead predictive of y_t:

$$y_t|D_{t-1} \sim T_{\theta n_{t-1}}(f_{t,}, q_t),$$

where

$$f_t = X_t' m_{t-1},$$

and

$$q_t = X_t' R_t X_t' + s_{t-1},$$

$$e_t = y_t - f_t,$$

The posterior for $\beta_t|v_t, D_t$ and $v_t|D_t$ can be now be calculated. The one-step-ahead predictive densities are calculated in the forward filter portion of FFBS, so model comparison can be conducted without doing backward sampling (or smoother) portion.