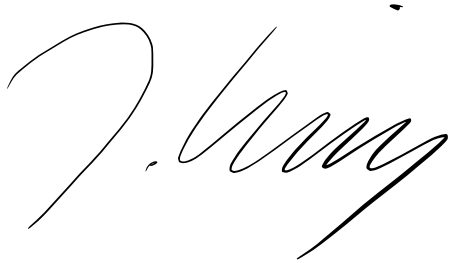
DISCLOSURE STATEMENT for
"Local Projections Are VAR Predictions of Increasing Order"

Julian F. Ludwig has no known conflicts of interest.

He is not receiving research funding from any organization for this research project.

He has no income related to the subject matter of this paper.

Nor do any of his relatives.



LOCAL PROJECTIONS ARE VAR PREDICTIONS OF INCREASING ORDER

JULIAN F. LUDWIG
Department of Economics, Texas Tech University
July 1, 2024

This paper explores local projections (LPs) and vector autoregressions (VARs), two important time series models in economics. I prove that LPs can be represented as combinations of equal and higher-order VARs, while VARs themselves are combinations of equal and lower-order LPs. This finding advances our understanding of LPs and VARs and leads to enhanced efficiency and model selection. Key contributions include: (i) enhancing model selection by comparing LPs and VARs that are equally parsimonious; (ii) improving the efficiency of LP estimation by leveraging the underlying VAR structure; and (iii) introducing the VAR sequence (VAR-S), a new empirical model that unifies LPs and VARs, thereby stimulating new avenues for research and advancements in the field.

KEYWORDS: Impulse Response Function, Multistep Forecasting, Local Projection, Vector Autoregression, VAR Sequence.

1. INTRODUCTION

Understanding the dynamic relationships among economic variables and their responses to various shocks is a fundamental goal in macroeconomic analysis. Central to this endeavor are methods for accurately modeling and forecasting these interactions. Vector autoregressions (VARs), popularized by Sims (1980), Bernanke (1986), and Blanchard and Watson (1986), have long been a cornerstone in capturing such dynamics. More recently, local projections (LPs), pioneered by Jordà (2005), have emerged as a direct approach for estimating impulse response functions (IRFs) and forecasts.[1] Despite their widespread application, the relationship between VARs and LPs has remained ambiguous, leading to ongoing debates in empirical research regarding model selection.[2] This paper aims to rec-

Julian F. Ludwig: Julian.Ludwig@ttu.edu

[1]For a comprehensive overview of the diverse applications of VAR and LP, see Ramey (2016), Nakamura and Steinsson (2018), and Plagborg-Møller and Wolf (2021a).

[2]For a debate between LP and VAR in IRF estimation, see Kilian and Kim (2011), Brugnolini (2018), Choi and Chudik (2019), Austin (2020), Bruns and Lütkepohl (2022), Plagborg-Møller and Wolf (2021a), and Li et al.

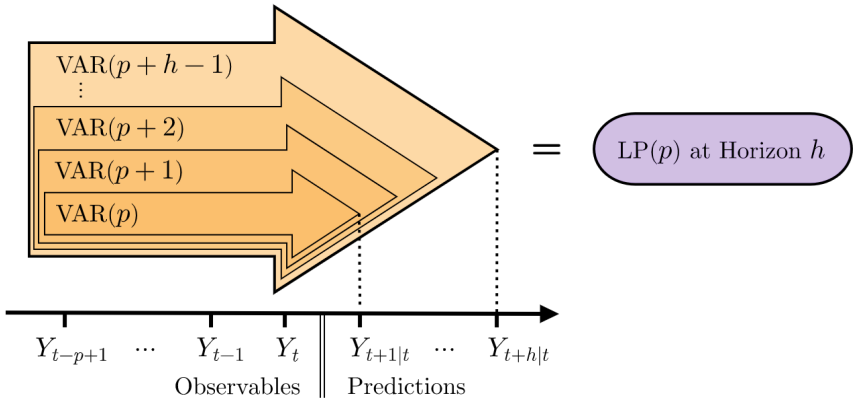


FIGURE 1.—Local Projections Are VAR Predictions of Increasing Order (Theorem 1)

Note: This diagram visualizes Theorem 1. It distinguishes between vector autoregressions and local projections, denoted as VAR(p) and LP(p), where p is the lag order of the model. Theorem 1 asserts that h-step ahead predictions made by LP(p) are equivalent to predictions generated by a sequence of h one-step ahead predictions made by VARs of increasing order. Specifically, a VAR(p) generates the first one-step ahead predictions, then a VAR($p+1$) advances the predictions by one period to two-step ahead predictions, then a VAR($p+2$) shifts the two-step ahead to three-step ahead predictions, and so on, until a VAR($p+h-1$) shifts the $(h-1)$-step ahead predictions to h-step ahead predictions, which, as stated by Theorem 1, are identical to the h-step ahead predictions made by an LP(p).

tify this gap in the literature by establishing a formal equivalence between LPs and VARs, thereby providing new insights and improving the efficiency and accuracy of macroeconomic modeling and forecasting.

The main finding of this paper is that local projections (LPs) and vector autoregressions (VARs) can be represented by each other. This equivalence is formalized as Theorem 1, which states that LPs are VAR predictions of increasing order, as depicted in Figure 1. The figure illustrates that an LP consists of a series of one-step-ahead predictions made by VARs, represented by arrows, with the number of lags increasing at each step. Conversely, the theorem also states that VARs are made up of LPs of decreasing order. This equivalence holds for both types of LPs: those estimated for a system of equations (Jordà, 2005), and the single-equation LPs that control for contemporaneous variables (Auerbach and Gorodnichenko, 2013a). Remarkably, Theorem 1 applies not only in population but also in sample when models are estimated with finite data. Therefore, Theorem 1 essentially introduces an

(2023); for a debate between direct forecasting (analogous to LP) and iterated forecasting (analogous to VAR) performance, see Schorfheide (2005), Chevillon (2007), Marcellino et al. (2006), Pesaran et al. (2011), and McElroy (2015).

alternative way to compute the exact same estimates, ensuring that any implications drawn from this theorem also apply in small sample settings.

This mathematical link between LPs and VARs brings a wealth of benefits for estimation, inference, identification, and selection of models when forecasting and computing impulse response functions (IRFs). This paper explores two immediate implications: Corollary 1 and 2, and three important applications of Theorem 1.

Corollary 1 states that since LPs consist of equal- and higher-order VARs, they are a generalization of an equal-order VAR. This is surprising, as they estimate the same number of parameters and thus appear to be equally restrictive. However, Theorem 1 reveals that by transforming LP parameters into VAR parameters, LPs implicitly estimate more VAR parameters than equal-order VARs. This explains the bias-variance tradeoff reported in the literature comparing LPs and VARs of equal order: more general models (LPs) are less biased but produce a larger sample variance than their special cases (equal-order VARs).

Corollary 2 shows that despite LPs being larger and more general than VARs, LP forecasts and IRFs still include equal-order VAR parameters at all horizons. Thus, an LP model can only be unbiased if the VAR of equal order is unbiased as well. However, since the VAR of equal order is more parsimonious, it becomes a better choice than an unbiased LP, making it illogical to estimate an unbiased LP. Therefore, the literature may benefit from moving away from the assumption of unbiasedness when estimating LPs. Since asymptotic theory often relies on models being unbiased, one way to address this is by assuming that the lag order increases with the sample size, thereby allowing the LP to be biased in finite samples but unbiased asymptotically (e.g., see Lewis and Reinsel, 1985, for an autoregressive model with an increasing lag order consistently estimating an infinite autoregression).

Now we can turn our attention to the three applications of Theorem 1. The first application fundamentally changes how we compare and select between LP and VAR models. By expressing LPs as VARs of increasing order, Theorem 1 allows for an "apples to apples" comparison of VAR to VAR parameters, rather than an "apples to oranges" comparison of VAR to LP parameters. With this in mind, I introduce the notion of model "size" as the number of VAR parameters used in a prediction. Comparing models of equal "size" ensures a fair comparison of equally parsimonious models, whereas comparing models of equal "order" pits a more general LP against a smaller VAR. The tradition of comparing LPs and VARs of equal "order" rather than "size" has led to a misleading bias-variance

tradeoff. While a bias-variance tradeoff might exist between smaller and larger models, there is no evidence of such a tradeoff between VARs and LPs once size is controlled for. Consequently, the literature may have incorrectly concluded that it is necessary to switch from a VAR to an LP if bias is prioritized, rather than switching from a smaller to a larger VAR. This highlights the need for updated research using size as the fixed measure rather than order.

The second application employs Theorem 1 to improve LP estimation. According to the theorem, the VARs that constitute an LP must be estimated using only a subset of the available data for the equivalence to hold. This unlocks the potential to increase the sample size. Specifically, when estimating forecasts or IRFs at horizon h, the new LP estimation method increases the average sample size by $\frac{1}{2}(h-1)$ periods, leading to improved performance in forecasting and IRF estimation, especially for large horizons.

The third application is the introduction of the VAR sequence (VAR-S), a novel empirical model that integrates LPs and VARs into a unified framework. While LPs and VARs restrict themselves to using either VARs of increasing or constant order for IRF and forecast estimation, the VAR-S offers more flexibility by allowing the choice of any sequence of VAR orders. This versatility opens up avenues for more accurate impulse response and forecast estimation, offering a promising new direction for the field. The VAR-S also enables the application of VAR tools to LPs, including the incorporation of error correction terms to capture cointegrating relationships (Engle and Granger, 1987), and the identification of IRFs using long-run restrictions (Blanchard and Quah, 1989). Additionally, because the VAR-S is more general, any new estimation, inference, or identification techniques developed for the VAR sequence will seamlessly apply to LPs and VARs, eliminating the need to explore these topics separately for each model.

In summary, this paper establishes a mathematical link between LPs and VARs, providing a deeper understanding of these models and paving the way for substantial breakthroughs in the field.

Literature

This study is situated within the literature on forecast and impulse response estimation. It contributes not only to the literature comparing models and improving model selection but also to the separate areas of research on local projections (or direct multistep forecasting)

and vector autoregressions (or iterated multistep forecasting).[3] That's because the new link between LPs and VARs not only allows for a better comparison but also enables the transfer of tools and insights from one model to the other, enhancing estimation, inference, and identification in both areas.

This paper addresses six specific findings in the literature, which are worth highlighting. First, Plagborg-Møller and Wolf (2021a) show that under certain assumptions, LPs and VARs of equal order produce the same impulse responses in population if either the lag order goes to infinity or if the IRF horizon does not exceed the lag order. Theorem 1 advances this research by deriving a precise link between LPs and VARs, not only in population but also in sample, and without imposing any restrictions on the data generating process. Essentially, the exact same results can be produced by either estimating an LP or its underlying VARs of equal and higher order.

Second, both the literature on model selection for impulse response estimation and the literature on multistep forecasting consistently observe a bias-variance tradeoff between LPs and VARs (see, for example, Pesaran et al., 2011, p. 174). This tradeoff suggests that LPs are more robust to misspecification, while VARs are more efficient. However, the underlying reason for this tradeoff has remained unclear.[4] Theorem 1 and its corollaries provide a clear explanation, revealing that LPs are more general and larger than they appear. This means that the observed bias-variance tradeoff may be an artifact of comparing larger LPs to smaller VARs, rather than an inherent property of the models themselves. [5]

---

[3] For an overview of comparisons between LP and VAR impulse response estimation, see Ramey (2016), Kilian and Lütkepohl (2017), Nakamura and Steinsson (2018), Stock and Watson (2018), and Plagborg-Møller and Wolf (2021a). For an overview of the literature comparing direct multistep forecasting (analogous to LP) with iterated multistep forecasting (analogous to VAR), see Bhansali (1999), Schorfheide (2005), Chevillon (2007), Marcellino et al. (2006), Pesaran et al. (2011), and McElroy (2015), with the debate tracing back to Cox (1961), who posits that direct methods could offer efficiency gains over the iterated approach.

[4] Svetunkov et al. (2023) observe that, even though there is a general understanding that direct multistep estimators lead to more conservative and robust models than iterated ones, as reported in studies like Cox (1961), Gersch and Kitagawa (1983), Tiao and Xu (1993), Marcellino et al. (2006), and McElroy (2015), these studies often fail to provide a clear explanation for this phenomenon.

[5] Since Plagborg-Møller and Wolf (2021a) find that LPs and VARs estimate the same thing as lag length goes to infinity, they reach the same conclusion that the bias-variance tradeoff is not inherent to the models. However, their findings do not apply to finite sample settings. Therefore, despite their asymptotic findings, the authors reaffirm that in finite samples and with finite lag lengths, researchers must navigate a bias-variance tradeoff between LPs and VARs (see Plagborg-Møller and Wolf, 2021a, p. 975). In contrast, Theorem 1 reveals that even in finite sample settings, the bias-variance tradeoff is an artifact of comparing larger LPs to smaller VARs. Consequently,

Third, the same literature often provides guidelines for applied researchers, such as recommending the choice of an LP over a VAR if unbiasedness is preferred over efficiency, or suggesting direct over iterated forecasting if the process is close to a unit root (Weiss, 1991, Liu, 1996), contains structural breaks (Tiao and Tsay, 1994), has a deterministic misspecification (Chevillon and Hendry, 2005), or has a misspecified moving average component (Schorfheide, 2005).[6] Moreover, specific characteristics of each model are noted, such as IRFs being more erratic and spiky when estimated with an LP than with a VAR (Ramey, 2016, Barnichon and Brownlees, 2019), or LPs underestimating IRFs in small sample settings (Gorodnichenko and Lee, 2020). However, most of these findings are derived by comparing LPs and VARs of the same order, thereby comparing a more general model to a smaller one, according to Theorem 1, and thus likely reflect differences in model size rather than model type. This paper converts LP parameters to VAR parameters via Theorem 1 to define model size as the number of VAR parameters, allowing for a comparison of equally parsimonious models. This approach avoids conflating model size and type, helping researchers select the right model for their research.

Fourth, the multistep forecasting literature highlights a major drawback of direct forecasting methods, such as local projection: the reduction in sample size resulting from including the forecast period in the estimation. In contrast, iterated forecasts maintain a larger sample size by limiting the gap between the dependent and explanatory variables to a single period (see, for example, Pesaran et al., 2011, p. 173). This issue is particularly relevant when the forecast horizon is large, consuming a substantial portion of the sample. Theorem 1 solves this problem by reinterpreting direct forecasts as iterated ones. Specifically, for LPs, this involves estimating the underlying iterated models—VARs of increasing

---

researchers may need to navigate a bias-variance tradeoff between larger and smaller models but not necessarily between LPs and VARs.

[6]For instance, a typical paper in this field investigates the models and derives practical conclusions by either performing and evaluating out-of-sample forecasts or by simulating data using various data generating processes (DGPs), then evaluating the accuracy of estimated IRFs or forecasts against the true ones generated by these DGPs. Through this lens, studies like Jordà (2005), Kilian and Kim (2011), Brugnolini (2018), Choi and Chudik (2019), Austin (2020), Bruns and Lütkepohl (2022), and Li et al. (2023) investigate the performance of LPs and VARs with respect to computing impulse responses. Similarly, Findley (1983), Findley (1985), Liu (1996), Kang (2003), Ang et al. (2006), Marcellino et al. (2006), and Pesaran et al. (2011) analyze the forecasting performance of direct and iterated multistep forecasting.

order—thereby making direct and iterated forecasting models equally efficient in their use of sample periods.

Fifth, since LPs and VARs are often used for similar purposes, several studies focus on adapting methods developed for VARs to local projections and vice versa. For instance, Nusbaum (2020) imposes long-run restrictions to identify IRFs in a local projection setting by assuming the IRFs eventually level out, while Gorodnichenko and Lee (2020) and Plagborg-Møller and Wolf (2022) introduce forecast error variance decompositions (FEVDs) to local projections, tools originally exclusive to VARs. The VAR sequence proposed here generalizes both LPs and VARs. Thus, any method developed for this new empirical model simultaneously applies to its special cases, LPs and VARs. Hence, the VAR sequence serves as an attractive baseline for introducing new methods and facilitates the transfer of tools between LPs and VARs. I demonstrate this by using a VAR sequence to estimate LP IRFs while still allowing for the identification of IRFs via long-run restrictions (Shapiro and Watson, 1988, Blanchard and Quah, 1989, King et al., 1991), as well as accommodating cointegrating relationships and trends (Engle and Granger, 1987, Johansen, 1991), tools commonly used in VARs but new to LPs.

Finally, Chang and Sakata (2007) and Plagborg-Møller and Wolf (2021a) establish equivalence in population between impulse responses estimated using vector LPs (with IRFs identified via a structural VAR) and single-equation LPs (with IRFs identified as regression coefficients or via instrument). I demonstrate that this connection also extends to sample estimates, provided that the VAR used for identifying the impulse has one fewer lag than the LP used to estimate the response. This lag order difference enables the two models to produce identical IRFs in both population and sample. This equivalence is formally stated in Lemma 1, a crucial step in applying Theorem 1 to all types of LPs, connecting them to VARs of equal and higher order. Moreover, Lemma 1 is significant in its own right, unifying the two modeling approaches so that all estimation and inference techniques developed for vector LPs are directly applicable to single-equation LPs, eliminating the need for separate methodologies.

Overview

The paper is structured as follows: Section 2 presents Theorem 1, demonstrating that LPs are composed of VARs of increasing order, and VARs are composed of LPs of de-

creasing order. The subsequent sections explore three applications: Section 3 introduces a new approach to selecting between LP and VAR models, Section 4 improves the efficiency of LP estimation, and Section 5 introduces the VAR sequence, a new empirical model that generalizes LPs and VARs. The paper concludes with Section 6. Some content is relegated to the appendix, including a generalized version of Theorem 1 to incorporate deterministic and stochastic trends in Appendix A. Additional details on the empirical exercises are presented in Appendix B, and while short versions of the proofs are included in the main text, more detailed proofs can be found in Appendix C.

2. LPS ARE VAR PREDICTIONS OF INCREASING ORDER

In this section, I present Theorem 1, which demonstrates that local projections (LPs) are composed of equal and higher-order vector autoregressions (VARs), leading to the conclusion that VARs are made up of equal and lower-order LPs. However, before presenting the theorem, it's important to establish the equivalence between the different types of LPs.

Section 2.1 defines the vector and single-equation local projection, along with the VAR. Section 2.2 outlines the assumptions required for estimation. Section 2.3 demonstrates the equivalence between vector and single-equation LP as Lemma 1. Section 2.4 introduces the concepts of VAR-based LPs and LP-based VARs. Section 2.5 establishes the equivalence between the models discussed in Section 2.1 and 2.4 as Theorem 1. Finally, Sections 2.6 and 2.7 present corollaries revealing that LPs generalize VARs of the same order (Corollary 1) and that LP estimation only makes sense when the model is biased (Corollary 2).

2.1. *Definition of LP, SELP, and VAR*

Local Projection (LP)

The local projection method of order p is denoted as LP(p) and projects future variables onto the linear space generated by current and past variables:[7]

[7]Local projection (LP) models utilize a technique in which endogenous variables, shifted several steps into the future, are regressed directly on explanatory variables. This method is employed for generating impulse response functions (IRFs) and for forecasting. The term "local projection" and its use for IRF estimation as an alternative to the vector autoregression (VAR) were introduced by Jordà (2005). Within the forecasting literature, LPs represent a linear approach to direct multi-step forecasting, also known as adaptive forecasting or dynamic estimation. The foundation of direct forecasting was set forth by Cox (1961), and Weiss (1991) established the consistency and

$$Y_{t+h} = b^{(h)}(p) + B_1^{(h)}(p)Y_t + \cdots + B_p^{(h)}(p)Y_{t-p+1} + r_{t+h}^{(h)}(p) \qquad h = 1, \ldots, H \tag{1}$$

where $\{Y_t\}_{t\in\mathbb{Z}}$ is an unknown stochastic process of N variables, $b^{(h)}(p)$ is an $N \times 1$ vector of constants, and $B_l^{(h)}(p)$ is an $N \times N$ matrix of coefficients for each lag l and horizon h. By definition of a linear projection, the residuals $r_{t+h}^{(h)}(p)$ have a mean of zero and are not correlated with regressors, i.e. $E\left[r_{t+h}^{(h)}(p)\right] = 0$, and $E\left[r_{t+h}^{(h)}(p)Y_{t+1-l}'\right] = 0$, for $l = 1, \ldots, p$.

The multistep forecast of a local projection follows directly from (1):

$$Y_{t+h|t} = b^{(h)}(p) + B_1^{(h)}(p)Y_t + \cdots + B_p^{(h)}(p)Y_{t-p+1} \qquad h = 1, \ldots, H \tag{2}$$

which then implies the following impulse response functions (IRFs):

$$IRF(h, d) = B_1^{(h)}(p)d \qquad h = 0, \ldots, H \tag{3}$$

with normalization $B_1^{(0)}(p) = I_N$, where I_N is an $N \times N$ identity matrix, and d is an $N \times 1$ vector that contains the contemporaneous effects of a shock of interest.

The impulse vector d in (3) is assumed to be known, but it can be estimated under additional assumptions. It is common to estimate d with a VAR, regardless of whether the IRFs are estimated with a local projection or a VAR (Jordà, 2005, 2009, Chang and Sakata, 2007, see). Thus, the estimation of d does not cause differences in the IRFs between the two models. An exception occurs when d is identified using restrictions on the estimated IRFs at $h > 0$. However, in such cases, if the model coefficients were the same, the IRFs and therefore d would also be the same. Hence, since Theorem 1 establishes an equivalence between model coefficients, the result does not hinge on whether the impulse vector d is known or estimated.

---

asymptotic normality of such forecasts under general conditions. The model termed local projection in this paper had been used as a direct forecasting method before it was termed local projection by Jordà (2005); for examples, see Tsay (1993), Lin and Tsay (1996), Bhansali (1996, 1997), Ing (2003). Although LP methods are linear in their parameters, they can model nonlinear relationships. This is achieved through a Taylor series expansion by including powers of the variables as regressors, as detailed by Jordà (2005, p. 167-168), or by employing regime-switching parameters, as explored by Auerbach and Gorodnichenko 2013a, 2013b, 2016.

Single-Equation Local Projection (SELP)

The single-equation local projection method of order p, denoted as SELP(p), differs from the system LP approach (1) by its inclusion of contemporaneous controls. Specifically, it projects future scalar variables onto the linear space spanned by current and past variables:[8]

$$\begin{aligned} y_{t+h} = \alpha^{(h)}(p) + \beta^{(h)}(p)x_t + \delta^{(h)}(p)'V_t \\ + \lambda_1^{(h)}(p)'Y_{t-1} + \cdots + \lambda_p^{(h)}(p)'Y_{t-p} + \nu_{t+h}^{(h)}(p) \quad h = 0, \ldots, H \end{aligned} \tag{4}$$

where y_t is the dependent scalar variable, x_t is the explanatory scalar variable, V_t is a $(K-1) \times 1$ vector of instant-effect control variables, and $Y_t = \left[V_t' \; x_t \; y_t \; W_t'\right]'$ is an $N \times 1$ vector that includes all variables, along with an $(N-K-1) \times 1$ vector of delayed-effect control variables W_t. The parameters $\alpha^{(h)}(p)$ and $\beta^{(h)}(p)$ are scalars, $\delta^{(h)}(p)$ is a $(K-1) \times 1$ vector of constants, and $\lambda_l^{(h)}(p)$ are $N \times 1$ vectors of constants for each lag $l \geq 1$ and horizon $h \geq 0$. The residuals $\nu_{t+h}^{(h)}(p)$, by definition of a linear projection, have a mean of zero and are uncorrelated with the regressors, i.e., $E\left[\nu_{t+h}^{(h)}(p)\right] = 0$, $E\left[\nu_{t+h}^{(h)}(p)x_t\right] = 0$, $E\left[\nu_{t+h}^{(h)}(p)V_t\right] = 0$, and $E\left[\nu_{t+h}^{(h)}(p)Y_{t-l}\right] = 0$ for $l = 1, \ldots, p$. It is noteworthy that the regressor x_t may be obtained from a first-stage regression using an instrumental variable, or it could be a shock identified in a separate model.

The SELP impulse response functions (IRFs) are derived from equation (4):[9]

---

[8] Focusing solely on a single equation in (1) simplifies the representation, estimation, and interpretation of local projections, as extensively practiced following the analyses of fiscal shocks by Auerbach and Gorodnichenko 2013a, 2013b, 2016 and the examination of oil shocks by Hamilton (2011). This simplicity has spurred various extensions, including the incorporation of contemporaneous regressors (Auerbach and Gorodnichenko, 2013a, Jordà et al., 2013, Favara and Imbs, 2015), the use of panel data (Auerbach and Gorodnichenko, 2013b, Owyang et al., 2013, Jordà et al., 2015, Jordà and Taylor, 2016), regime-switching parameters (Auerbach and Gorodnichenko, 2013a,b, 2016, Hamilton, 2011, Tenreyro and Thwaites, 2016), the inclusion of an exogenous shock identified in a separate model (Chang and Sakata, 2007, Ramey and Zubairy, 2018), and instrumental variable (IV) local projections (Jordà et al., 2015, 2020, Leduc and Wilson, 2017, Ramey and Zubairy, 2018, Stock and Watson, 2018). For comprehensive reviews on local projections' diverse applications, see Ramey (2016), Nakamura and Steinsson (2018), Adämmer (2019), Plagborg-Møller and Wolf (2021a).

[9] When the dependent variable is differenced, the cumulated impulse response function (CIRF) is typically obtained by summing the IRF parameters as outlined in equation (refeq:selpirf). However, within the SELP literature, it is a common practice to cumulate the dependent variable first and then directly estimate the cumulated IRF (Favara and Imbs, 2015, Auerbach and Gorodnichenko, 2016, Jordà and Taylor, 2016, e.g., see). This approach streamlines the inference process. However, due to the linear properties of the covariance, i.e. $\text{Cov}\left(\sum_{s=0}^{h} y_{t+s}, x_t\right) = \sum_{s=0}^{h} \text{Cov}(y_{t+s}, x_t)$, the resulting CIRFs are identical to summing up the IRF parameters in (5).

$$IRF(h) = \beta^{(h)}(p) \qquad h = 0, \ldots, H \tag{5}$$

where $\beta^{(h)}(p)$ is the scalar coefficient associated with the explanatory variable x_t in (4). It's important to note that SELP restricts some of the contemporaneous relationships. Specifically, the effects of x_t and V_t are controlled for, while changes in y_t and W_t are not. Therefore, the estimated coefficient $\beta^{(h)}(p)$ measures the impact of x_t on y_{t+h}, holding V_t constant, but allowing W_t to react to x_t. The SELP IRF thus reflects the response to an impulse affecting all variables except W_t.

Vector Autoregression (VAR)

The vector autoregression (VAR) of order p is denoted as VAR(p) and projects current variables onto the linear space generated by past variables:[10]

$$Y_t = \gamma(p) + \Gamma_1(p)Y_{t-1} + \cdots + \Gamma_p(p)Y_{t-p} + \xi_t(p) \tag{6}$$

where $\gamma(p)$ is an $N \times 1$ vector of constants, and $\Gamma_l(p)$ is an $N \times N$ matrix of coefficients for each lag l. By definition of a regression, the VAR residuals $\xi_t(p)$ have a mean of zero and are not correlated with regressors, i.e. $E[\xi_t(p)] = 0$, and $E[\xi_t(p)Y'_{t-l}] = 0$, for $l = 1, \ldots, p$. It is common to assume that VAR residuals are not correlated over time, however, VAR estimation does not enforce this assumption. As a result, the residuals may actually be correlated in practice, meaning that $E[\xi_t(p)\xi_{t-k}(p)'] \neq 0$ is possible for some $k \neq 0$.

In contrast to local projections, VARs are not capable of directly estimating IRFs and forecasts. Instead, they require an iterative process. The derivation of IRFs and forecasts is

[10] Before local projection methods were introduced, the primary approach for multivariate IRF and forecast estimation was the vector autoregression (VAR), a framework established by Sims (1980). The VAR relates each variable to its own lagged values and those of other variables within the system. To produce forecasts and IRFs, VARs iteratively apply the estimated relationships from one time period to the next. Thus, VARs represent a linear approach to iterated multi-step forecasting, in contrast to the direct multi-step forecasting approach of local projections. For a comprehensive overview on VAR's diverse applications, see Hendry (1995), Johansen (1995), Hamilton (1994), Banerjee et al. (1993), Pfaff (2008), Ramey (2016), Nakamura and Steinsson (2018), Plagborg-Møller and Wolf (2021a). A related method is the autoregressive distributed lag (ADL) model, which is a restricted form of the VAR and serves as a single-equation alternative for estimating IRFs. Romer and Romer (2004, 2010) use this approach by including an exogenous shock as one of the variables, ensuring it does not appear on the left-hand side. For an overview of ADLs, see Baek and Lee (2022).

achieved in three steps, which are outlined below. Although these steps may be obvious to some readers, I still outline them because in Section 2.4, I demonstrate how to modify these steps to develop a VAR equivalent of a local projection. The first step involves duplicating the same VAR h times, while shifting the time index incrementally:

$$\begin{aligned} Y_{t+1} &= \gamma(p) + \Gamma_1(p)Y_t + \cdots + \Gamma_p(p)Y_{t+1-p} + \xi_{t+1}(p) \\ Y_{t+2} &= \gamma(p) + \Gamma_1(p)Y_{t+1} + \cdots + \Gamma_p(p)Y_{t+2-p} + \xi_{t+2}(p) \\ &\vdots \\ Y_{t+h} &= \gamma(p) + \Gamma_1(p)Y_{t+h-1} + \cdots + \Gamma_p(p)Y_{t+h-p} + \xi_{t+h}(p) \qquad h = 1, \ldots, H \end{aligned} \tag{7}$$

The second step entails combining the duplicated VARs to substitute all regressors corresponding to time periods later than t:

$$Y_{t+h} = \pi^{(h)}(p) + \Pi_1^{(h)}(p)Y_t + \cdots + \Pi_p^{(h)}(p)Y_{t-p+1} + \zeta_{t+h}^{(h)}(p) \qquad h = 1, \ldots, H \tag{8}$$

where coefficients and forecast errors are the following functions of the VARs in (7):

$$\begin{aligned} \Pi_l^{(h)}(p) &= \sum_{k=1}^{p} \Gamma_k(p)\Pi_l^{(h-k)}(p) && l = 1, \ldots, p \\ \pi^{(h)}(p) &= \gamma(p) + \sum_{k=1}^{\min\{h-1,p\}} \Gamma_k(p)\pi^{(h-k)}(p) \\ \zeta_{t+h}^{(h)}(p) &= \xi_{t+h}(p) + \sum_{k=1}^{\min\{h-1,p\}} \Gamma_k(p)\zeta_{t+h-k}^{(h-k)}(p) \qquad h = 1, \ldots, H \end{aligned} \tag{9}$$

with normalization $\Pi_l^{(1-k)}(p) = I_N$ for all $l = k \geq 1$, and $\Pi_l^{(1-k)}(p) = 0$ otherwise.

The third and final step involves computing the multistep forecast by removing the forecast errors in equation (8):

$$Y_{t+h|t} = \pi^{(h)}(p) + \Pi_1^{(h)}(p)Y_t + \cdots + \Pi_p^{(h)}(p)Y_{t-p+1} \qquad h = 1, \ldots, H \tag{10}$$

which then implies the following impulse response functions (IRFs):

$$IRF(h,d) = \Pi_1^{(h)}(p)d \qquad h = 0,\dots,H \tag{11}$$

with normalization $\Pi_1^{(0)}(p) = I_N$, where h is the IRF horizon, d is the impact of the shock, and p is the lag order.

Typically, the impulse vector d in (11) is obtained by decomposing the VAR residuals into uncorrelated shocks in what is known as a structural vector autoregression (SVAR), as pioneered by Bernanke (1986), Blanchard and Watson (1986), Blanchard and Quah (1989), Sims and Zha (1998), Uhlig (2005). Nonetheless, as discussed earlier, Theorem 1 applies regardless of whether d is assumed to be known or estimated, because, for a given identification strategy, identical model coefficients lead to the estimation of identical impulse vectors.

To summarize, local projections estimate forecasts and impulse responses directly, while VARs require an iterative process to derive them. Another difference is that the forecast errors of a local projection are not correlated with the variables used in the forecast, while the forecast errors of a VAR prediction in (8) may be correlated with some of the right-hand side variables, i.e. $E\left[\zeta_{t+h}^{(h)}(p)Y'_{t+1-l}\right] \neq 0$, for some $h \geq 2$, and $l \geq p-(h-2)$, is possible. This is because LP forecast errors and LP residuals are the same thing, which by construction, are uncorrelated with the regressors.

2.2. *Estimation*

The models are estimated with a time series sample $\{y_t\}_{t\in\{1,\dots,T\}}$ generated by an unknown stochastic process $\{Y_t\}_{t\in\mathbb{Z}}$, by replacing the orthogonality conditions in the models with sample estimates, as detailed in Appendix C.[11] Lemma 1 and Theorem 1 do not impose any restrictions on the stationarity or ergodicity of the stochastic process. The only requirement is that the time series has finite second moments, in order to ensure that all models are well-defined. Hence, the estimates don't need to be unique, unbiased, or con-

[11] Since Lemma 1 and Theorem 1 show equivalence not only in population but also in sample, the sample periods included in the estimation need to match as well. Thus, the detailed versions of the proofs of Lemma 1 and Theorem 1 in Appendix C refine the definitions of the models by stating the sample used for estimation and the respective OLS formulas.

sistent, as the aim is not to connect them to population parameters, but to relate the sample estimates of one model to another.

2.3. *Lemma 1: Vector LPs Replicate Single-Equation LPs*

To make the comparison between LPs and VARs easier, I show that the impulse response functions (IRFs) from a single equation local projection (SELP) in (5) are the same as those from a vector local projection (LP) in (3) for a certain impulse vector d. This way, the rest of the paper can just focus on vector LPs, knowing that they give the same results as SELPs.

The comparison between the LP in (1) and the SELP in (4) reveals that LP regressors cover p periods, whereas SELP regressors cover $p+1$ periods. This necessitates the use of an LP($p+1$) model to replicate the SELP(p) model. In addition, the impulse vector d of the LP IRFs in (3) must be zero for the variables that are controlled for in the SELP, particularly for the first $K-1$ elements in Y_t that correspond to the instant-effect control variables V_t. The remaining components of d are determined using the residuals from a VAR(p) model. Notably, the VAR model has one lag less than the LP($p+1$) model because the VAR(p) residuals, influenced by both dependent and explanatory variables, span $p+1$ periods, which aligns with the $p+1$ periods covered by the SELP(p) regressors.

The impulse vector d for replicating the IRF of an SELP(p) with an LP($p+1$) is derived by decomposing the covariance matrix of the VAR(p) residuals $\xi_t(p)$ in (1) as follows:

$$d = \frac{\mathbf{A}e_K}{e_K'\mathbf{A}e_K}, \qquad \mathbf{A} = \mathrm{Chol}\Big(\mathrm{Cov}\big(\xi_t\left(p\right)\big)\Big) \tag{12}$$

Here, e_K is an N-dimensional vector consisting of all zeros except for a one in the Kth position. This vector is used to select the Kth column from the preceding matrix or the Kth row from the succeeding matrix. $\mathrm{Cov}\left(\cdot\right)$ computes the covariance matrix, either for the sample or the population, and $\mathrm{Chol}\left(\cdot\right)$ performs Cholesky decomposition to produce a lower-triangular matrix. The numerator of the impulse vector d is a vector with its first $(K-1)$ elements being zero due to the lower-triangular nature of $\mathbf{A}$. This aligns with the SELP(p)'s inclusion of V_t as control variables, implying that the first $K-1$ elements of Y_t cannot respond on impact. The denominator of d normalizes the vector so that its Kth element equals one, representing a one-unit increase in the Kth variable x_t of the LP($p+1$)

model. This is consistent with SELP(p)'s IRF measuring the impact of a one-unit increase in x_t.

The Frisch-Waugh-Lovell (FWL) Theorem transforms the SELP into a combination of LP and VAR coefficients, thereby showing that SELP(p) and LP($p+1$) produce the same IRFs when the impulse vector is derived with a VAR(p):

LEMMA 1—Vector LPs Replicate Single-Equation LPs: *The SELP(p) IRFs in (5) are equivalent to those from the LP($p+1$) for the $(K+1)$th variable. This equivalence holds when generating LP($p+1$) IRFs in (3) using the VAR(p) impulse vector from (A.14). This holds true in both population and sample, implying that the IRFs remain identical even when the model parameters and residuals are replaced by their sample estimates.*

PROOF: This equivalence is derived by applying the FWL Theorem, stated as Lemma B.1 in Appendix C.1, to transform the SELP(p) in (4) into a combination of LP($p+1$) and VAR(p) in (1) and (6). See Appendix C.2 for a detailed proof. *Q.E.D.*

Lemma 1 demonstrates that vector local projections (LPs) can precisely emulate single-equation local projections (SELPs). This extends the work of Chang and Sakata (2007) and Plagborg-Møller and Wolf (2021a,b), who showed such equivalence in population under certain assumptions. Lemma 1 achieves equivalence not just in population but also in sample by removing one of the lags of the structural VAR. This finding is significant because it unifies two different modeling methods into one approach. Thus, the focus can be on vector LPs, recognizing that SELPs are special cases within this broader framework. Consequently, all estimation and inference techniques used for LPs are also applicable to SELPs, eliminating the need for separate methodologies. Additionally, it is straightforward to demonstrate that instrumental-variable SELPs can be replicated using vector LPs by scaling the impulse vector d accordingly. Therefore, this type of model is also encompassed within the LP framework of expressions (1) and (3).

2.4. *Definition of VAR-Based LP and LP-Based VAR*

The goal is to represent local projections as combinations of VARs and vice versa. This is achieved by increasing the lag order of the duplicated VARs in equation (7) incrementally, such that the forecast errors remain uncorrelated with the variables used in the forecast,

mimicking the characteristics of a local projection. Theorem 1 confirms that the VAR predictions produced in this manner are equivalent to local projections. This, in turn, implies that VARs can be expressed as LPs of decreasing order.

VAR Version of a Local Projection

To obtain a VAR version of a local projection, I follow the same three steps for deriving VAR forecasts outlined in Section 2.1. However, in the first step, instead of duplicating the same VAR h times, I increase the number of lags by one at every iteration:

$$\begin{aligned}
Y_{t+1} &= \gamma(p) + \Gamma_1(p) Y_t + \cdots + \Gamma_p(p) Y_{t+1-p} + \xi_{t+1}(p) \\
Y_{t+2} &= \gamma(p+1) + \Gamma_1(p+1) Y_{t+1} + \cdots + \Gamma_{p+1}(p+1) Y_{t+1-p} + \xi_{t+2}(p+1) \\
&\vdots \\
Y_{t+h} &= \gamma(p+h-1) + \Gamma_1(p+h-1) Y_{t+h-1} \\
&\quad + \cdots + \Gamma_{p+h-1}(p+h-1) Y_{t+1-p} + \xi_{t+h}(p+h-1) \qquad h = 1, \ldots, H
\end{aligned} \tag{13}$$

where $\gamma(p+h-1)$, $\Gamma_l(p+h-1)$, and $\xi_{t+k}(p+h-1)$ are VAR($p+h-1$) coefficients and residuals in (6). Because of the increasing lag order, none of the VAR residuals are correlated with LP(p) regressors, i.e. $E\left[\xi_{t+h}(p+h-1) Y'_{t+1-l}\right] = 0$, for $l = 1, \ldots, p$.

In the second step, the VARs are combined to replace future with past variables:

$$Y_{t+h} = \psi^{(h)}(p) + \Psi_1^{(h)}(p) Y_t + \cdots + \Psi_p^{(h)}(p) Y_{t-p+1} + \chi_{t+h}^{(h)}(p) \qquad h = 1, \ldots, H \tag{14}$$

where coefficients and forecast errors are the following functions of the VARs in (13):

$$\begin{aligned}
\Psi_l^{(h)}(p) &= \sum_{k=1}^{h+l-1} \Gamma_k(p+h-1) \Psi_l^{(h-k)}(p) \qquad l = 1, \ldots, p+h-1 \\
\psi^{(h)}(p) &= \gamma(p+h-1) + \sum_{k=1}^{h-1} \Gamma_k(p+h-1) \psi^{(h-k)}(p) \\
\chi_{t+h}^{(h)}(p) &= \xi_{t+h}(p+h-1) + \sum_{k=1}^{h-1} \Gamma_k(p+h-1) \chi_{t+h-k}^{(h-k)}(p) \qquad h = 1, \ldots, H
\end{aligned} \tag{15}$$

with normalization $\Psi_l^{(1-k)}(p) = I_N$ for all $l = k \geq 1$, and $\Psi_l^{(1-k)}(p) = 0$ otherwise. Since the VAR residuals in (13) are uncorrelated with LP(p) regressors, so are the residuals in (14), i.e. $E\left[\chi_{t+h}^{(h)}(p) Y'_{t+1-l}\right] = 0$, for $l = 1, \ldots, p$. Therefore, the VAR-based predictions in (14) meet the same orthogonality conditions as a local projection in (1).

The final step is to obtain multistep forecasts by removing the forecast errors in (14):

$$Y_{t+h|t} = \psi^{(h)}(p) + \Psi_1^{(h)}(p) Y_t + \cdots + \Psi_p^{(h)}(p) Y_{t-p+1} \qquad h = 1, \ldots, H \tag{16}$$

which then implies the following impulse response functions (IRFs):

$$IRF(h, d) = \Psi_1^{(h)}(p) d \qquad h = 0, \ldots, H \tag{17}$$

with normalization $\Psi_1^{(0)}(p) = I_N$, where h is the IRF horizon, d is the impact of the shock, and p is the lag order of the initial VAR.

To summarize, by increasing the number of lags by one at every itaration, the VAR predictions in (14) have the same form and properties as a local projection (1), and it is therefore not surprising that they are equivalent, which is the first statement of Theorem 1. Before presenting the theorem, let's switch perspective and write a VAR in terms of local projections.

Local Projection Version of a VAR

To derive a local projection version of a VAR, we must ensure that the LPs only depend on variables that are considered in the VAR. This is achieved by combining all LPs that include as many variables as possible but no more than the VAR(p) itself, which consists of $\{Y_t, \ldots, Y_{t-p}\}$:

$$\begin{aligned}
Y_t &= b^{(1)}(p) + B_k^{(1)}(p) Y_{t-1} + \cdots + B_p^{(1)}(p) Y_{t-p} + r_t^{(1)}(p) \\
Y_t &= b^{(2)}(p-1) + B_1^{(2)}(p-1) Y_{t-2} + \cdots + B_{p-1}^{(2)}(p-1) Y_{t-p} + r_t^{(2)}(p-1) \\
&\vdots \\
Y_t &= b^{(h)}(p-h+1) + B_1^{(h)}(p-h+1) Y_{t-h} \\
&\quad + \cdots + B_{p-h+1}^{(h)}(p-h+1) Y_{t-p} + r_t^{(h)}(p-h+1) \qquad h = 1, \ldots, p+1
\end{aligned} \tag{18}$$

where $b^{(h)}(p-h+1)$, $B_l^{(h)}(p-h+1)$, and $r_t^{(h)}(p-h+1)$ are LP$(p-h+1)$ coefficients and residuals at horizon h, as defined in expression (1). Note that the number of LP lags does not exceed p in order to prevent dependence on variables outside the VAR(p). For example, the local projection of Y_t on $\{Y_{t-h},\ldots,Y_{t-p}\}$ only has $p-(h-1)$ lags so that all regressors beyond lag p are excluded.

The conversion of (19) to an LP-based VAR model occurs in two steps. In the first step, the VAR predictions (8) are substituted with LPs of decreasing order (18). In the second step, the predictions of the first step are combined according to (9) to form an LP-based VAR model:

$$Y_t = a(p) + A_1(p)Y_{t-1} + \cdots + A_p(p)Y_{t-p} + v_t(p) \tag{19}$$

where the coefficients and residuals are determined by the LPs in (18):

$$\begin{aligned}
A_l(p) &= B_{l-h+1}^{(h)}(p-h+1) - \sum_{k=1}^{h-1}\Gamma_k(p)\,B_{l-h+1}^{(h-k)}(p-h+1) \qquad l=h,\ldots,p \\
a(p) &= b^{(h)}(p-h+1) - \sum_{k=1}^{h-1}\Gamma_k(p)\,b^{(h-k)}(p-h+1) \\
v_t(p) &= r_t^{(h)}(p-h+1) - \sum_{k=1}^{h-1}\Gamma_k(p)\,r_{t-k}^{(h-k)}(p-h+1) \qquad h=1,\ldots,p+1
\end{aligned} \tag{20}$$

Note that the LP-based VAR model in (19) does not depend on forecast horizon h. As a result, the coefficients and residuals in (20) can be obtained for any value of $h=1,\ldots,p+1$, providing multiple ways to derive the same solution. Nevertheless, regardless of the choice of h, the coefficients and residuals will always be derived from LPs of at most order p.

To summarize, by decreasing the number of lags when increasing the local projection horizon, the LP-based VAR model (19) depends on the same variables as the VAR(p). Hence, it is not surprising that they are equivalent, which is the second statement of Theorem 1.

2.5. *Theorem 1: LPs Are Combinations of VARs and Vice Versa*

The FWL Theorem is used to transform the LP and VAR models of Section 2.1 into their respective VAR and LP counterparts of Section 2.4, thereby establishing equivalence:

THEOREM 1—Local Projections Are Combinations of VARs and Vice Versa: *The following holds true in both population and sample:*

(a) A local projection is a combination of equal and higher-order VARs, and

(b) a VAR is a combination of equal and lower-order local projections.

Specifically, (a) expressions (1) and (14) are equivalent, and (b) expressions (6) and (19) are equivalent, even when parameters and residuals are replaced by their sample estimates.

PROOF: Statements (a) and (b) are derived by applying the FWL Theorem to transform the LP and VAR models in (1) and (6) into their respective VAR and LP counterparts in (14) and (19). See Appendix C.3 for a detailed proof. *Q.E.D.*

Theorem 1 connects LPs and VARs by showing they can be represented by each other. This breakthrough enhances our understanding of both models, resulting in advancements in model selection, efficiency, and inference. The remainder of the paper harnesses Theorem 1 to improve the model selection process (Section 2.6, 2.7, and 3), enhance the efficiency of LP estimation (Section 4), and introduce the VAR sequence (VAR-S), a model that generalizes LPs and VARs (Section 5). However, before delving into these applications, let's illustrate Theorem 1 with an example and relate it to the findings of Plagborg-Møller and Wolf (2021a).

Example

The purpose of this example is to show that Theorem 1 is not just a theoretical result but holds in practice when the models are estimated with finite data. The data and underlying theory for this example can be found in Appendix B.

Figure 2 plots IRFs and forecasts of LPs and VARs (1) and (6) and their VAR- and LP-counterparts (14) and (19), estimated with U.S. macro indicators. As Theorem 1 (a) states, LP-IRFs and forecasts are equal to the ones produced by VARs of increasing order. This is evident as the bright dashed lines of the VAR-based LP model (14) match the dark thick lines of the LP model (1). Additionally, in accordance with Theorem 1 (b), VAR-IRFs and

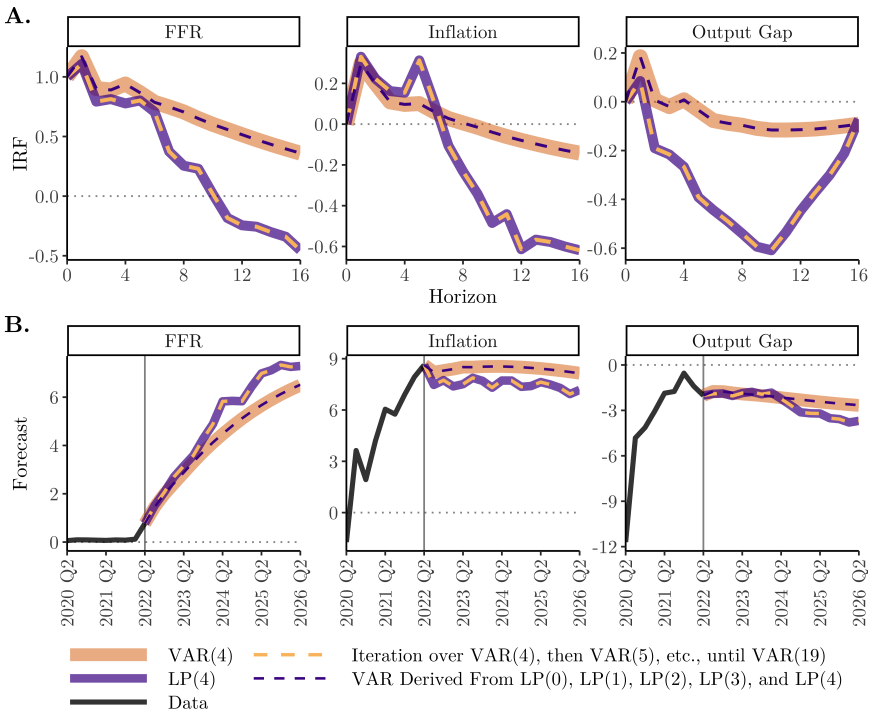


FIGURE 2.—Estimated LPs as Combinations of VARs and Vice Versa (Theorem 1)

Note: All variables are in annual percentage rates. The figure illustrates Theorem 1, which states that LPs are combinations of equal and higher-order VARs, and VARs are combinations of equal and lower-order LPs. The example uses U.S. macro indicators from 1955 Q1 to 2021 Q1, as plotted in Figure B.2. Panel A displays the impulse response functions (IRFs) to a one percentage point increase in the federal funds rate (FFR), while holding current and past variables constant. Panel B shows the forecasts for the next sixteen quarters. Four models are used: (i) VAR(4) as defined in (8), (ii) LP(4) as defined in (1), (iii) an iteration over VAR(4), then VAR(5), etc., until VAR(19) as defined in (14), and (iv) a VAR derived from LP(0), LP(1), LP(2), LP(3), and LP(4) models as defined in (19). As stated in Theorem 1, the IRFs and forecasts of (i) and (iv) are identical, as are those of (ii) and (iii). The estimation of the models is done by OLS, and can be found in Appendix C. The data and underlying theory for this example, a replication of Jordà (2005) with updated data, are described in Appendix B.

forecasts are equivalent to the ones produced by LPs of decreasing order. This is shown as the dark dashed lines of the LP-based VAR model (19) overlap the bright thick lines of the VAR model (6). Thus, Figure 2 illustrates that Theorem 1 remains valid even when the models are estimated using finite data.

Generalization of Plagborg-Møller and Wolf (2021)

In a recent paper, Plagborg-Møller and Wolf (2021a) show that LP(p) and VAR(p) models produce the same IRFs in population when lag length goes to infinity, i.e. $p \to \infty$ (Proposition 1), or as long as the IRF horizon h does not exceed lag length, i.e. $h \leq p$ (Proposition 2). Theorem 1 allows to summarize these claims in a single expression; specif-

ically, Plagborg-Møller and Wolf (2021a) claim the following:

$$B_1^{(h)}(p)d - \Pi_1^{(h)}(p)d = \sum_{k=1}^{h}\sum_{s=1}^{k} \Pi_1^{(h-k)}(p) \underbrace{\left[\Gamma_s(p+k-1) - \Gamma_s(p)\right]}_{=0} B_1^{(k-s)}(p)d \qquad (21)$$
$$h = 1, \ldots, p$$

where $B_1^{(h)}(p)d$ and $\Pi_1^{(h)}(p)d$ are the LP- and VAR-IRFs of (3) and (11), $\Gamma_s(r)$ are the VAR(r) coefficients defined in (6), and therefore, the terms in brackets are VAR lag order selection biases. The normalizations are $\Pi_1^{(0)}(p) = B_1^{(0)}(p) = I_N$, $B_1^{(1-s)}(p) = 0$ for $s > 1$, and $\Gamma_s(r) = 0$ for $s > r$. This expression reveals that Proposition 1 requires VARs to be convergent so that the terms inside the brackets go to zero as $p \to \infty$, and Proposition 2 requires the VAR(p) coefficients to remain unchanged when adding up to $p-1$ lags to the model.

Theorem 1 broadens the two propositions in Plagborg-Møller and Wolf (2021a) in five dimensions. Firstly, the assumptions needed for the propositions to be true are disclosed in an analytical expression relating the two models, rather than being imposed in advance. Secondly, expression (21) is applicable not only in population but also in sample, allowing for analysis in small sample settings. Thirdly, Theorem 1 allows for comparisons of not only IRFs but also forecasts. Fourthly, Theorem 1 enables comparisons between LPs and VARs of distinct orders, not just those of equal order. This becomes important in Section 3, where I argue that instead of comparing LPs and VARs of the same order, a more meaningful comparison can be made by comparing LPs and VARs of the same size, as measured by the number of underlying VAR parameters. Lastly, the equivalence in Theorem 1 also holds when incorporating deterministic trends and cointegrating relationships, as shown in Theorem A.1 of Appendix A, thereby generalizing the two propositions to vector error correction models (VECMs).

2.6. *Corollary 1: LPs Generalize VARs and Vice Versa*

Theorem 1 states that a local projection is composed of VARs of equal and higher order, making it both a generalization and a special case of a VAR:

COROLLARY 1—LPs Generalize VARs and Vice Versa:

(a) LP(p) generalize VAR(p) predictions, and

(b) VAR($p+H-1$) generalize LP(p) predictions, at horizon $h=1,\ldots,H$.

PROOF: Theorem 1 demonstrates that LP(p) is constructed from VARs of equal and higher order, making it a generalization of VAR(p), thus proving (a). Expression (13) and Theorem 1 show that VAR($p+H-1$) encompasses all the VARs that form LP(p) at horizons $h=1,\ldots,H$, thereby generalizing LP(p), thus proving (b). *Q.E.D.*

Corollary 1 reveals that LPs and VARs can be both special cases and generalizations of each other for certain model orders. Hence, they estimate the same data generating process and cannot capture any aspects of the process that the other model cannot.

Interestingly, statement (a) of Corollary 1 reveals that comparing LPs with VARs of equal order, a common practice in the literature, effectively pits a more general model against a more specific one. This is akin to comparing a VAR with more lags against a VAR with fewer lags. This explains the bias-variance tradeoff observed in the literature, where LPs exhibit lower bias at the expense of higher estimation variance compared to VARs. Corollary 1 shows that this bias-variance tradeoff happens not because LPs handle bias better, but because the comparisons are made between LPs and VARs of the same order, making the LP a generalization of the VAR.[12]

As a result, if the results from an LP contradict those from a VAR of the same order, the VAR results can be dismissed in favor of the LP results. This makes model selection quite simple and uninteresting. Therefore, instead of comparing LPs and VARs of equal *order*, Section 3 introduces the concept of comparing LPs and VARs of equal *size*, defined as models using the same number of VAR parameters after transforming the LP into VAR parameters via Theorem 1. This way, if one model is better at reducing bias, it's because of its inherent structure, and not because it is a generalization of the other.

[12] The insight that the bias-variance tradeoff is not intrinsic to LPs and VARs was first made by Plagborg-Møller and Wolf (2021a), who show that LPs and VARs estimate the same thing in an asymptotic environment where sample size and model size go to infinity. They conclude that VAR impulse response estimators are not generally more efficient than local projection estimators, nor are local projections generally more robust to misspecification than VARs. However, their finding does not translate to finite sample settings, and consequently, they acknowledge that in finite samples and with finite lag lengths, researchers must navigate a bias-variance tradeoff between LPs and VARs (see Plagborg-Møller and Wolf, 2021a, p. 975).

2.7. *Corollary 2: LPs Have to Be Biased*

Theorem 1 and Corollary 1 imply that it only makes sense to estimate a local projection if the model is biased:

COROLLARY 2—Unbiased LPs Are Inefficient: *An unbiased LP(p) is always inferior to a VAR(p).*

PROOF: Theorem 1 and Corollary 1 imply that the LP(p) model is biased if and only if the VAR(p) model is biased. The "if" portion stems from Theorem 1, which states that LP(p) is a combination of equal and higher-order VARs. This means that LP(p) forecasts and IRFs always incorporate VAR(p) parameters, allowing any bias in the VAR(p) to transfer to the LP(p). Conversely, the "only if" portion is derived from Corollary 1, which shows that the LP(p) model is a generalization of the VAR(p) model. This implies that if the LP(p) model is biased, then so are its special cases, including the VAR(p). Furthermore, Theorem 1 reveals that an LP(p) model consists of equal and higher-order VAR(q) models, where $q \geq p$. This means it estimates more coefficients than a VAR(p), making it less efficient if the VAR(p) is unbiased. Hence, when the LP(p) model is unbiased, both models are well-specified, but the VAR(p) is more efficient and thus superior to the LP(p). Because Theorem 1 holds both in sample and population, the statement is always true, even in small sample settings. *Q.E.D.*

Corollary 2 reveals that when the LP(p) model is unbiased, it is better to estimate a VAR(p). Therefore, when estimating a local projection, one cannot impose the convenient assumption of unbiasedness, as this would make the VAR a better choice. Instead, one should assume that IRFs and forecasts are biased regardless of the model, and therefore, the goal is to minimize the bias, not eliminate it. Bhansali (1996, 1999) describe this approach as a non-parametric procedure, approximating an unknown data generating process (DGP) instead of establishing what the true DGP is.[13]

[13]By allowing the DGP to differ from the model, it is common to let the model that approximates the DGP, often referred to as the forecast-generating process (FGP), have its number of parameters increase with sample size, thereby getting larger and potentially closer to the DGP as the sample size increases (see Bhansali, 1999, for further details).

Allowing LPs to be biased presents a challenge for estimation, inference, and identification methods that rely on correctly specified models. For instance, it is common practice to identify the impulse of an IRF by decomposing VAR residuals into economic shocks. But this leads to a dilemma: allowing a biased model means that the shocks are no longer identified, while claiming the shocks are identified contradicts the estimation of a local projection instead of a VAR. A possible solution is to assume that the lag order increases with the sample size, allowing LPs to be biased in finite samples but unbiased asymptotically. This approach aligns with the work of Lewis and Reinsel (1985) on the consistency of models with increasing lag orders and Inoue and Kilian (2002) on the asymptotic validity of the bootstrap method when the underlying process is an infinite autoregression.

Corollary 2 confirms a finding consistently observed in the literature on direct multistep forecasting (analogous to LP) versus iterated multistep forecasting (analogous to VAR): direct forecasting is inefficient for well-specified models (see Chevillon, 2007, Section 3). Theorem 1 enriches the understanding of this issue in two key ways. First, it provides a clear reason for this inefficiency: a local projection is less efficient than a VAR of the same order because it uses a larger number of estimated VAR parameters. Second, while Johnston (1974) and Kabaila (1981) show that a well-specified direct forecasting model is less efficient than the iterated counterpart, they derive this result only asymptotically. Because they restrict themselves to an asymptotic setting, they conclude that gains of direct forecasting may still be found in small samples, even if the model is well-specified. Theorem 1, which applies to both asymptotic settings and sample estimates, allows Corollary 2 to contradict this by stating that unbiased LPs are always inferior to VARs, even in small samples.

3. APPLICATION I: A NEW MEASURE OF MODEL SIZE

In this section, I use Theorem 1 to argue against comparing LPs and VARs of the same *order*, suggesting instead to compare them based on *size*. Section 3.1 defines models of equal *size* as those using the same number of VAR parameters for predictions, while Section 3.2 highlights how comparing LPs and VARs of the same order (rather than size) has led to misconceptions about a bias-variance tradeoff in the literature.

3.1. *Definition of Model Size*

Although LPs and VARs of equal order have an equal number of parameters, Theorem 1 shows that LPs are larger as they consist of higher-order VARs. The theorem further reveals that LPs and VARs share a common denominator in VAR parameters, which calls for the following apples to apples measure of model size:

DEFINITION 1—Model Size: The size of a model is the average number of VAR parameters used per iteration and equation to forecast a single variable.

(i) A VAR forecast employs the same VAR for every iteration (7), hence, the size of an h-step ahead VAR(q) prediction with N variables and a constant is

$$\text{size}\left\{\text{VAR}(q), h\right\} = \frac{N(1+Nq)+\cdots+N(1+Nq)}{hN} = 1+Nq.$$

(ii) According to Theorem 1, an LP uses VARs of increasing order (13), therefore, the size of an h-step ahead LP(p) prediction with N variables and a constant is

$$\text{size}\left\{\text{LP}(p), h\right\} = \frac{N(1+Np)+\cdots+N(1+N(p+h-1))}{hN} = 1+N\left(p+\frac{1}{2}(h-1)\right).$$

In summary, the size of a model refers to the average number of estimated coefficients per regression used to iterate a variable h steps forward. Therefore, a smaller model is more parsimonious than a larger one.[14]

Inspection of Definition 1 reveals the following:

COROLLARY 3—LPs and VARs of the Same Order vs. Size: *For any horizon* $h = 1, 2, 3, \ldots$ *and* N *variables,*

(a) LPs are $50\frac{(h-1)}{(p+\frac{1}{N})}\%$ *larger than VARs of the same order* p*, and*

(b) LPs and VARs are of the same size if the VAR model has $\frac{1}{2}(h-1)$ *more lags.*

[14]While LPs and VARs of equal size are equally parsimonious, LPs implicitly estimate VAR parameters less efficiently due to losing more observations than VARs. To address this data loss issue, Section 4 introduces a new LP estimation method that explicitly estimates VARs before aggregating them into LPs via Theorem 1. This approach ensures that models of equal size are not only equally parsimonious but also equally efficient in terms of sample utilization.

PROOF: (a) An LP(p) at horizon h is $X\%$ larger than a VAR(p) if

$$X = 100\left(\frac{\text{size}\left\{\text{LP}(p), h\right\} - \text{size}\left\{\text{VAR}(p), h\right\}}{\text{size}\left\{\text{VAR}(p), h\right\}}\right) = 50\frac{(h-1)}{\left(p + \frac{1}{N}\right)} \qquad h = 1, 2, 3, \ldots$$

where model size is defined in Definition 1. (b) LP(p) and VAR(q) models are of the same size if

$$\begin{aligned} \text{size}\left\{\text{VAR}(q), h\right\} &= \text{size}\left\{\text{LP}(p), h\right\} \\ 1 + Nq &= 1 + N\left(p + \frac{1}{2}(h-1)\right) \\ q &= p + \frac{1}{2}(h-1) \qquad\qquad h = 1, 3, 5, \ldots \end{aligned}$$

and therefore, the VAR lag order q needs to be $\frac{1}{2}(h-1)$ periods larger to match the LP lag order p. *Q.E.D.*

An example of Corollary 3 (a) is that an LP(4) is at least 50% larger than a VAR(4) at horizon $h = 6$, which means that the LP model uses 50% more VAR parameters to make a six-step ahead prediction. Moreover, an LP(15) is at least 15% larger than a VAR(15) at horizon $h = 6$, and an LP(4) is at least 140% larger than a VAR(4) at horizon $h = 15$. An example of Corollary 3 (b) is that an LP(4) is as large as a VAR(11) at horizon $h = 15$.

Corollary 3 along with Corollary 1 provide a straightforward explanation for the well-established result that local projections are more robust and less efficient than VARs of the same order: the tradeoff between bias and variance occurs simply because LPs are larger! Even though they estimate the same number of LP or VAR parameters, LPs implicitly estimate more VAR parameters than VARs, making them less parsimonious. Since the size of an LP model increases with forecast horizon h, while the size of a VAR remains constant, this tradeoff becomes more pronounced for larger horizons.

Note that Corollaries 1 and 3 suggest that adding $\frac{1}{2}(h-1)$ lags to either an LP or a VAR of the same size at horizon $h = 1, 3, 5, \ldots$ makes that model a generalization of the other. Conversely, removing that amount of lags makes that model a special case of the other. Hence, LPs and VARs of equal size are not only equally parsimonious but also equally restrictive.

3.2. *Dispelling the Myth of a Bias-Variance Tradeoff*

LPs and VARs of equal order may seem equally parsimonious as they estimate the same number of LP and VAR parameters. However, Corollary 3 reveals that this is a result of an "apples to oranges" comparison. Once the LP parameters are transformed to VAR parameters via Theorem 1, an "apples to apples" comparison reveals that LPs are larger than VARs of equal order. Therefore, a fairer comparison is between LPs and VARs of equal size, which keeps the number of VAR parameters fixed. For example, if an LP is less biased than a VAR of the same size in a specific scenario, it can be concluded that the LP uses its parameters more efficiently under those conditions. Such conclusion would not be valid if the models were of the same order, as it would not be clear whether the reduction in bias was due to the model structure or simply the result of the LP estimating more VAR parameters.

Comparing LPs and VARs of equal order rather than size may have led much of the literature astray. While a bias-variance tradeoff exists between smaller and larger models, there is no evidence that such a tradeoff exists between LPs and VARs once size is controlled for. The purported bias-variance tradeoff may have misled researchers who prioritize unbiasedness over efficiency to switch from a VAR to an LP, when the bias reduction might have been greater by switching to a lager VAR. Hence, research that finds such a tradeoff may need to be revisited, investigating whether the tradeoff stems from comparing larger to smaller models or if the bias reduction is indeed an inherent property of local projections within the specific context of the comparison. This reevaluation will provide a more accurate understanding of the comparative advantages of these models, thereby improving model selection.

To summarize, I dispel the "myth" that LPs are generally more robust to misspecification and VARs are more efficient. While Plagborg-Møller and Wolf (2021a) have already dispelled this myth in an asymptotic context, their findings do not apply to finite sample settings. Therefore, despite their asymptotic findings, the authors end up reaffirming the claim that in finite samples and with finite lag lengths, researchers must navigate a bias-variance tradeoff between LPs and VARs (see Plagborg-Møller and Wolf, 2021a, p. 975). However, Theorem 1, together with Corollaries 1 and 3, reveals that this tradeoff may be

an artifact of comparing a more general LP to a smaller VAR in the literature, rather than an inherent property of the models in finite samples.

4. APPLICATION II: A NEW LP ESTIMATION METHOD

This section utilizes Theorem 1, which reveals that LPs are composed of VARs of increasing order, to enhance the efficiency of local projection estimation. The approach involves estimating and assembling the underlying VARs rather than directly estimating the LP, thereby increasing the sample size. While it is already established that estimating a local projection by horizon $h = 1, \dots, H$ results in a gain of $H - h$ observations compared to joint estimation, the new method gains an additional $\frac{1}{2}(h-1)$ observations, improving the efficiency of local projections particularly for long horizons.

Section 4.1 explains LP estimation, introduces the new VAR-based estimation method, and demonstrates that the new method increases sample size. Section 4.2 exemplifies the improved performance of the new method in predicting key macroeconomic indicators.

4.1. *Sample Size for Different LP Estimation Methods*

Estimation is based on a time series sample $\{y_t\}_{t\in\{1,\dots,T\}}$ with T periods drawn from a stochastic process $\{Y_t\}_{t\in\mathbb{Z}}$. The regressors are required to have full rank and all second moments must be finite in sample. As in Section 2, the goal is not to connect the estimates to population parameters but to compare sample sizes across estimation methods, and thus do not need to be unbiased or consistent.

Traditional LP Estimation Method

Traditionally, the local projection of order p, as defined in (1), is estimated using the following method,

$$\left[\hat{b}^{(h)}(p)\ \hat{B}_1^{(h)}(p) \cdots \hat{B}_p^{(h)}(p)\right] = \sum_{t\in\mathcal{S}^{(h)}} y_{t+h} x_t(p)' \left(\sum_{t\in\mathcal{S}^{(h)}} x_t(p) x_t(p)' \right)^{-1} \tag{22}$$
$$h = 1, \dots, H, \quad \mathcal{S}^{(h)} \subseteq \{p, \dots, T-h\}$$

where $x_t(p) = \left[1\, y_t' \cdots y_{t-p+1}'\right]'$ are sample regressors for the LP(p) model, $\hat{b}^{(h)}(p)$ and $\hat{B}_l^{(h)}(p)$ are the sample estimates of the LP parameters in (1), and $\mathcal{S}^{(h)}$ is the time series sample used for estimation.

There are two common ways to estimate LPs, (i) estimation *by maximum horizon*, also known as joint estimation, where the sample stays the same for all horizons, i.e. $\mathcal{S}^{(h)} = \{p, \ldots, T-H\}$, for all h, and (ii) estimation *by horizon*, where the sample size increases for shorter horizons, i.e. $\mathcal{S}^{(h)} = \{p, \ldots, T-h\}$. The third method, which I propose, is (iii) estimation *by VAR*, which increases the sample size even more.

New VAR-Based LP Estimation Method

The parameters of the VAR-based LP model (14) of order p at horizon h are estimated in two steps. In the first step, the underlying VARs (13) of increasing order: VAR($p+j-1$), for $j = 1, \ldots, h$, are estimated as follows:

$$\left[\hat{\gamma}(s_j)\, \hat{\Gamma}_1(s_j) \cdots \hat{\Gamma}_{s_j}(s_j)\right] = \sum_{t \in \mathcal{T}_h} y_t x_{t-1}(s_j)' \left(\sum_{t \in \mathcal{T}_h} x_{t-1}(s_j) x_{t-1}(s_j)' \right)^{-1}$$
$$s_j = p + j - 1, \quad j = 1, \ldots, h, \quad \mathcal{T}_j \subseteq \{p+j, \ldots, T\} \tag{23}$$

where $x_{t-1}(s_j) = \left[1\, y_{t-1}' \cdots y_{t-s_j}'\right]'$ are sample regressors for the VAR($p+j-1$), $\hat{\gamma}(s_j)$ and $\hat{\Gamma}_l(s_j)$ are the sample estimates of the VAR parameters in (6), and $\mathcal{T}_j$ is the time series sample used for VAR($p+j-1$) estimation.

In the second step, the VAR parameters estimated in (23) are assembled into LP coefficients according to expression (24):

$$\begin{aligned} \hat{\psi}^{(h)}(p) &= \hat{\gamma}(p+h-1) + \sum_{k=1}^{h-1} \hat{\Gamma}_k(p+h-1) \hat{\psi}^{(h-k)}(p) \\ \hat{\Psi}_l^{(h)}(p) &= \sum_{k=1}^{h+l-1} \hat{\Gamma}_k(p+h-1) \hat{\Psi}_l^{(h-k)}(p) \qquad\qquad l = 1, \ldots, p+h-1 \end{aligned} \tag{24}$$

with normalization $\hat{\Psi}_l^{(1-k)}(p) = I_N$ for all $l = k$, and $\hat{\Psi}_l^{(1-k)}(p) = 0$ otherwise, where $\hat{\psi}^{(h)}(p)$ and $\hat{\Psi}_l^{(h)}(p)$ are the sample estimates of the VAR-based LP parameters in (14).

Theorem 1 states that the LP coefficients in (22) are equal to the VAR-based LP coefficients in (24), both in population and sample, when using a specific time series sample of $\mathcal{T}_j = \{p+j, \ldots, T-H+j\}$. However, the sets of periods $\mathcal{T}_j$ used for VAR estimation in Theorem 1 allow for the possibility of including more time periods. In particular, the VAR($p+j-1$) in equation (23) can be estimated using a maximum of $T-(p+j-1)$ time periods if the sample used for estimation is $\mathcal{T}_j = \{p+j, \ldots, T\}$. The new VAR-based LP estimation method is defined as estimating (24) using this largest possible sample. It is referred to as LP estimation "*by VAR*" because the same VAR coefficients are used to calculate LPs at all horizons, unlike traditional estimation methods which use different VAR samples for each horizon, as shown in Corollary 4 below.

Sample Size Comparison

To show that the new VAR-based LP estimation method increases sample size, Corollary 4 lists the VAR samples that mimic traditional LP estimation methods, then compares the sample sizes between these and the new estimation method:

COROLLARY 4—VAR Samples by LP Estimation Method:
Theorem 1 applies if the time periods used for LP estimation $\mathcal{S}^{(h)}$ in (22) match the following time periods used for VAR estimation $\{\mathcal{T}_1, \ldots, \mathcal{T}_h\}$ in (23):

	LP(p) *at Horizon* h	*VAR*$(p+j-1)$, *for* $j=1,\ldots,h$
(i) *By Max Horizon*	$\mathcal{S}^{(h)} = \{p, \ldots, T-H\}$	$\mathcal{T}_j = \{p+j, \ldots, T-H+j\}$
(ii) *By Horizon*	$\mathcal{S}^{(h)} = \{p, \ldots, T-h\}$	$\mathcal{T}_j = \{p+j, \ldots, T-h+j\}$
(iii) *By VAR*	(*not possible*)	$\mathcal{T}_j = \{p+j, \ldots, T\}$

$$h = 1, \ldots, H$$

where method (iii) is a new way of estimating an LP(p) using all available periods. The last column reveals that the new method (iii) uses more data for VAR estimation than the traditional methods (i) and (ii).

PROOF: The proof of Theorem 1 in Appendix C.3 shows that (i) is true. Note that (i) implies (ii) for an LP(p) at horizon $h = H$; and therefore, by setting the maximum horizon to $H = 1$, $H = 2$, etc., (i) derives (ii) for all horizons. *Q.E.D.*

This corollary provides a more precise formulation of Theorem 1, as it shows exactly how the VARs need to be estimated so that the VAR-based LP (14) is equivalent to an LP (1) in sample. Moreover, Corollary 4 reveals that the traditional LP estimation methods remove more periods than necessary at the VAR-level; and therefore, LP estimation can be improved by including these periods.

Method (iii) in Corollary 4 makes two significant contributions. First, it improves efficiency in estimating IRFs and forecasts using LPs, contributing to the existing literature on local projection estimation. Second, it addresses the sample size problem that has been a concern in the literature on multistep forecasting (see, for example, Pesaran et al., 2011, p. 173). Direct forecasting methods, like LPs, include the forecasted period in the estimation, which reduces the sample size by h periods when estimating h-step ahead forecasts. In contrast, iterated forecasting methods, like VARs, maintain a larger sample size by having only a one-period gap between dependent and explanatory variables. This sample size problem worsens with longer forecast horizons, effectively shrinking to zero as the horizon increases. The new VAR-based LP estimation method solves this problem by estimating one-step models, using the same sample size as iterated models. This approach can also benefit other direct forecasting models by converting them into a series of one-step predictions before estimation.

Table I compares the VAR sample sizes for the three LP(p) estimation methods outlined in Corollary 4. The columns represent VAR models of increasing order, while the rows represent forecast horizons. Method (i) estimates an LP with the same sample for all horizons; and therefore, the sample sizes are identical across all VARs. Method (ii) employs different samples for different LP horizons, causing the sample size for each VAR to vary based on the horizon; thus, the VAR sample sizes are the same within rows, but differ across rows. Method (iii) utilizes the largest available sample for each VAR model, resulting in the same sample size within each column, but differing from column to column.

Moreover, Table I lists the average sample size for a VAR-based LP of order p both *at horizon* h and *until horizon* H. The former calculates the average size for a specific horizon h, while the latter calculates it for the entire IRF or forecast path until horizon H. At horizon h, Table I shows that estimation (ii) *by horizon* instead of (i) *by maximum horizon* increases the sample size by $H - h$ periods, and estimation (iii) *by VAR* instead of (ii) *by horizon* increases the sample again by $\frac{1}{2}(h-1)$ periods. While the gain in sample

TABLE I

SAMPLE SIZE BY LP ESTIMATION METHOD (COROLLARY 4)

LP Estimation Method		**VAR Sample Size**				**Average**
Horizon \ *VAR*		VAR(p)	VAR($p+1$)	$\cdots$	VAR($p+H-1$)	
(i) By Max Horizon	1	$T-p-(H-1)$				<u>At Horizon</u> h
	2					$T-p-(H-1)$
	$\vdots$					
	H					<u>Until Horizon</u> H
						$T-p-(H-1)$
(ii) By Horizon	1	$T-p$				<u>At Horizon</u> h
	2	$T-p-1$				$T-p-(h-1)$
	$\vdots$	$\vdots$				
	H	$T-p-(H-1)$				<u>Until Horizon</u> H
						$T-p-\frac{1}{2}(H-1)$
(iii) By VAR	1	$T-p$	$T-p-1$	$\cdots$	$T-p-(H-1)$	<u>At Horizon</u> h
	2					$T-p-\frac{1}{2}(h-1)$
	$\vdots$					<u>Until Horizon</u> H
	H					$T-p-\frac{1}{4}(H-1)$

Note: This table compares sample sizes for the three LP estimation methods of Corollary 4 with T time periods and p lags in the local projection. The columns depict VAR models of increasing order, VAR($p+j-1$), $j=1,\ldots,H$, while the rows depict forecast horizons, $h=1,\ldots,H$. The number in row "h" and column "VAR(k)" is the sample size of the VAR(k) used for the VAR-based LP at horizon h that replicates estimation (i) *by maximum horizon*, (ii) *by horizon*, or (iii) *by VAR*. Method (i) uses the same sample for all LPs, resulting in identical sample sizes for all VARs and horizons. Method (ii) uses different samples for different horizons, leading to varying sample sizes across rows, but with the same sample size within each row. Method (iii) maximizes the sample size for each VAR and then uses the same VAR for all horizons, causing sample sizes to vary across columns but remain the same within columns. The last column provides the average sample sizes for each of the three estimation methods. The average sample size <u>at horizon h</u> is the average across all columns of the hth row, and the average sample size <u>until horizon H</u> is the average across all columns and all rows. Formally, the former is defined as $\mu_h = \frac{1}{h}\sum_{j=1}^{h} \#\mathcal{T}_j$, where $\#\mathcal{T}_j$ is the number of elements in the VAR sample $\mathcal{T}_j$ of Corollary 4, and the latter is calculated as $\frac{1}{h}\sum_{k=1}^{h}\mu_k$.

size of (ii) is large for short horizons, (iii) also increases the sample size for long horizons; hence, the new method is particularly useful for estimating long-term IRFs and forecasts. For the entire IRF or forecast up to horizon H, Table I shows that method (ii) cuts the lost periods by half at the end of the sample compared to (i), and method (iii) cuts the lost periods by half again, on average, compared to (ii).

Let's explore the implications of increasing the sample size with an example.

4.2. *Improved Performance of New LP Estimation Method*

To compare performance, I calculate the out-of-sample mean squared errors (MSEs) of LP(4) forecasts when estimated (i) *by maximum horizon*, (ii) *by horizon*, and (iii) *by VAR*, as defined in Corollary 4. I compare the methods in their ability to forecast three U.S.

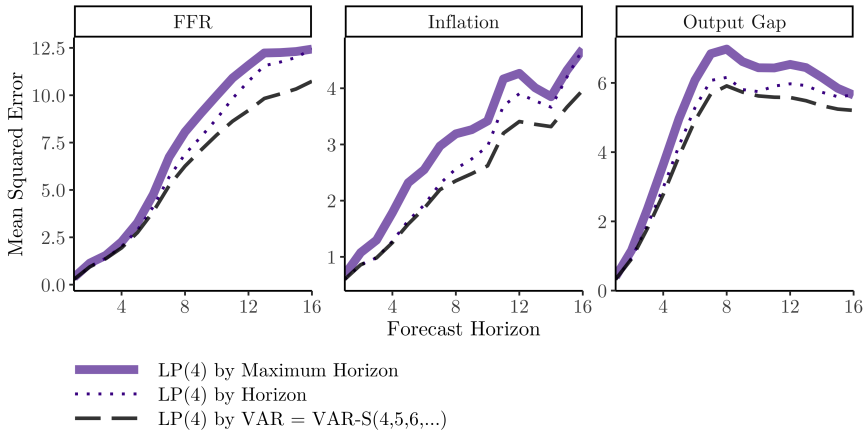


FIGURE 3.—Out-Of-Sample Performance by LP Estimation Method (Corollary 4)

Note: This figure plots the out-of-sample mean squared errors (MSEs) across forecast horizons for LP(4) forecasts estimated using three methods: estimation (i) *by maximum horizon*, (ii) *by horizon*, and (iii) *by VAR*, as defined in Corollary 4. The LP forecast estimated by VAR has the lowest MSE and thus the best out-of-sample performance, as the underlying VARs are estimated using all available periods, not just a subset, as seen in Corollary 4 and Table I. The MSE is calculated as the average of the squared forecast errors of 140 LP forecasts estimated using a rolling forecasting scheme. The scheme involves splitting 67 years of US macroeconomic indicators from 1955 Q1 to 2022 Q2 into 140 samples of 30 years (120 quarters) each, then using these samples to estimate the model and generate 140 out-of-sample forecasts. More details on the rolling forecasting scheme and data can be found in Appendix B.

macroeconomic indicators: federal funds rate (FFR), inflation, and output gap, which are the same variables used in Jordà (2005) to compare local projections to VARs. The MSE is calculated using a rolling forecasting scheme, where the quarterly data from 1955 Q1 to 2022 Q2 is split into 140 separate samples of 30 years. For each sample, an out-of-sample forecast is made, and the mean squared error (MSE) is determined by taking the average of the squared errors of those 140 forecasts. More details on the rolling forecasting scheme and data can be found in Appendix B.

Figure 3 displays the performance of the three LP estimation methods from Corollary 4. The plot shows that the LPs estimated (iii) *by VAR* have the lowest MSE across all horizons and thus perform better than LPs estimated (i) *by maximum horizon* and (ii) *by horizon*. This is because the VARs are estimated using all available periods, not just a subset, as demonstrated in Corollary 4 and Table I. Method (ii) outperforms method (i), but the advantage decreases with forecast horizon and eventually becomes negligible, while method (iii) consistently outperforms the other methods, especially at larger horizons.

To summarize, a new method of estimating local projections (LPs) has been developed that makes use of the insight that LPs are vector autoregressions (VARs) of increasing order

(Theorem 1). By estimating the underlying VARs, fewer time periods need to be removed from the sample, resulting in a larger sample size. This method increases forecasting performance, as demonstrated in a comparison of out-of-sample mean squared errors.

5. APPLICATION III: VAR SEQUENCE - A GENERALIZATION

In this section, a novel empirical model called VAR sequence (VAR-S) is introduced, which unifies local projections (LPs) and vector autoregressions (VARs). Theorem 1 highlights that LPs can be interpreted as a combination of VARs of increasing order, whereas VAR predictions are made up of VARs of constant order. By exploiting this observation, the VAR sequence enables the selection of an arbitrary sequence of VAR orders, rather than being restricted to an increasing or constant order.

This generalization offers three main benefits: First, it broadens the range of lag order options, potentially improving IRF and forecast estimation. Second, the VAR sequence can be designed to incorporate features from both LP and VAR methods, leveraging the strengths of each. This allows for the application of VAR tools to LPs, such as including error correction terms to capture cointegrating relationships as in Engle and Granger (1987) or identifying IRFs using long-run restrictions as in Blanchard and Quah (1989). Third, future research can apply new estimation, inference, or identification techniques to the VAR sequence, thereby improving both LP and VAR methods simultaneously, eliminating the need to explore these topics separately for each model.

Section 5.1 provides a detailed definition of the VAR sequence. Section 5.2 elaborates on how both LPs and VARs fit into the framework of VAR sequences and introduces additional hybrid types. Finally, Section 5.3 explores the use of the VAR sequence to apply VAR tools to LPs.

5.1. *Definition of VAR Sequence (VAR-S)*

A VAR sequence (VAR-S) is a method that estimates IRFs and forecasts based on a series of VARs. The VAR-S(s), defined by a particular *lag sequence* $s = \{s_1, \ldots, s_H\}$ or denoted as VAR-S($s_1, \ldots, s_H$), operates by estimating a VAR at each lag order $s_h \in s$. These individual VARs are then combined to produce IRFs and forecasts through a three-step procedure, akin to the procedure used to compute the VAR-based LP in Section 2.4.

In the first step, the VARs of order s_j, for $j = 1, \dots, h$, are listed in order, while shifting the time index incrementally:

$$\begin{aligned}
Y_{t+1} &= \gamma(s_1) + \Gamma_1(s_1)Y_t + \cdots + \Gamma_{s_1}(s_1)Y_{t+1-s_1} + \xi_{t+1}(s_1) \\
Y_{t+2} &= \gamma(s_2) + \Gamma_1(s_2)Y_{t+1} + \cdots + \Gamma_{s_2}(s_2)Y_{t+2-s_2} + \xi_{t+2}(s_2) \\
&\vdots \\
Y_{t+h} &= \gamma(s_h) + \Gamma_1(s_h)Y_{t+h-1} + \cdots + \Gamma_{s_h}(s_h)Y_{t+h-s_h} + \xi_{t+h}(s_h) \qquad h = 1, \dots, H
\end{aligned} \tag{25}$$

where $\gamma(s_h)$, $\Gamma_l(s_h)$, and $\xi_t(s_h)$ are the VAR(s_h) coefficients and residuals in (6).

In the second step, the VARs are combined to replace future regressors with current and past variables:

$$\begin{aligned}
Y_{t+h} = g^{(h)}(s) + G_1^{(h)}(s)Y_t + \cdots + G_{\lambda_s}^{(h)}(s)Y_{t-\lambda_s+1} + v_{t+h}^{(h)}(s) \\
h = 1, 2, \dots, \quad \lambda_s = \max_h\{s_h + 1 - h\}
\end{aligned} \tag{26}$$

where coefficients and forecast errors are the following functions of the VARs in (25):

$$\begin{aligned}
G_l^{(h)}(s) &= \sum_{k=1}^{s_h} \Gamma_k(s_h) G_l^{(h-k)}(s) && l = 1, \dots, \lambda_s \\
g^{(h)}(s) &= \gamma(s_h) + \sum_{k=1}^{\min\{h-1, s_h\}} \Gamma_k(s_h) g^{(h-k)}(s) \\
v_{t+h}^{(h)}(s) &= \xi_{t+h}(s_h) + \sum_{k=1}^{\min\{h-1, s_h\}} \Gamma_k(s_h) v_{t+h-k}^{(h-k)}(s) && h = 1, 2, \dots
\end{aligned} \tag{27}$$

with normalizations $G_l^{(1-k)}(s) = I_N$ for all $l = k \geq 1$, and $G_l^{(1-k)}(s) = 0$ otherwise.

The final step is to obtain multistep forecasts by removing the forecast errors in (26):

$$Y_{t+h|t} = g^{(h)}(s) + G_1^{(h)}(s)Y_t + \cdots + G_{\lambda_s}^{(h)}(s)Y_{t-\lambda_s+1} \qquad h = 1, \dots, H \tag{28}$$

which then implies the following impulse response functions (IRFs):

$$IRF(h,d) = G_1^{(h)}(s)d \qquad h = 0,\ldots,H \tag{29}$$

with normalization $G_1^{(0)}(s) = I_N$, where h is the IRF horizon, d is the impact of the shock, and $s = \{s_1,\ldots,s_H\}$ is the lag sequence.

5.2. *Types of VAR Sequences*

Since LPs are combinations of VARs of increasing order and VAR predictions are combinations of VARs of the same order, it follows that both LP and VAR predictions are types of VAR sequences:

COROLLARY 5—LPs and VARs Are VAR Sequences:
(a) A VAR(q) produces the same IRFs and forecasts as a VAR-S($q,q,q,\ldots$), and
(b) an LP(p) produces the same IRFs and forecasts as a VAR-S($p,p+1,p+2\ldots$).
This is true in both, population and sample.

PROOF: As per definition of a VAR sequence, the VAR-S($q,q,q,\ldots$) is a VAR(q) prediction model, as defined in equation (8), and the VAR-S($p,p+1,p+2\ldots$) is a VAR-based LP model, as defined in equation (14), which corresponds to a local projection (1), according to Theorem 1. Because Theorem 1 is true in both, population and sample, Corollary 5 is true in both, too. *Q.E.D.*

Corollary 5 shows that LPs and VARs are both instances of VAR sequences, where the lag sequence increases linearly for LPs and remains constant for VARs. However, it should be noted that the lag sequence does not need to adhere to either of these two patterns, and it is possible that superior patterns exist that have yet to be discovered.

VAR sequences allow for arbitrary combinations of LPs and VARs. For instance, since the lag sequence of an LP has a slope of one: $s = \{p,p+1,p+2,\ldots\}$, and the lag sequence of a VAR has a slope of zero: $s = \{p,p,p,\ldots\}$, a combination with an average slope of one half: $s = \{p,p,p+1,p+1,p+2,\ldots\}$ is possible. Alternatively, a VAR sequence with an average slope of $1/4$ moves it closer to a VAR, and an average slope of $3/4$ brings it closer

to an LP. Thus, the VAR-S framework allows for model selection on a spectrum, rather than a binary choice between LPs and VARs.

While more research needs to be done to uncover the best VAR sequences, I will describe three new types below that incorporate useful features of LPs and VARs.

DEFINITION 2—LP-VARs, VAR-LPs, and VAR-LP-VARs:
Define the following types of VAR-S(s) with lag sequence $s = \{s_h\}_{h\in\mathbb{N}}$:

(i) LP(p) = VAR-S($p, p+1, \ldots$), i.e. $s_h = p + h - 1$,

(ii) VAR(q) = VAR-S($q, q, \ldots$), i.e. $s_h = q$,

(iii) LP-VAR(p, q) = VAR-S($p, p+1, \ldots, q, q, \ldots$), i.e. $s_h = \min\{p + h - 1, q\}$,

(iv) VAR-LP(m, p) = VAR-S($m, \ldots, m, m+1, \ldots$), i.e. $s_h = \max\{m, p + h - 1\}$, and

(v) VAR-LP-VAR(m, p, q) = VAR-S($m, \ldots, m, m+1, \ldots, q, q, \ldots$), i.e. $s_h = \min\{\max\{m, p + h - 1\}, q\}$.

The LP-VAR(p, q) mimics an LP(p) but limits the VAR order to a maximum of q periods instead of letting it increase indefinitely. This enables predictions at long horizons, which a standard LP(p) cannot achieve as the model becomes too large for those horizons. To keep it similar to an LP, q should be chosen as large as possible but small enough for efficient VAR estimation. For instance, when $q = p + H - 1$, the LP-VAR(p, q) matches the predictions of an LP(p) up to horizon H, while still being able to make long-run predictions.

The LP-VAR(p, q) can also be viewed as mimicking VAR(q) predictions, but with the forecast depending only on p lags instead of q. Specifically, the forecast of Y_{t+h} will be a function of $\{Y_t, \ldots, Y_{t-p+1}\}$. This is useful, for instance, when there is a need to exclude recent events like the COVID-19 crisis from the forecast, even if the event occurred less than q periods ago.

The VAR-LP(m, p) acts like a VAR(m) but maintains the local projection property where forecast errors are orthogonal to regressors $\{Y_t, \ldots, Y_{t-p+1}\}$ across all horizons. This ensures there is no omitted variable bias caused by excluding those regressors in the IRFs and forecasts across all horizons. Therefore, estimating a VAR-LP(m, p) mitigates potential bias if the assumption that the data generating process (DGP) follows a VAR(m) is violated.

The VAR-LP-VAR(m, p, q) combines the strengths of all four models in Definition 2. When $p \leq m \leq q$, this model first uses VARs of order m for predictions, then incorporates

VARs of increasing order similar to an LP(p), and finally switches to VARs of order q once the horizon exceeds $q-p$. This flexibility allows VAR-LP-VAR models to compute long-run IRFs and forecasts while ensuring that forecast errors remain orthogonal to regressors $\{Y_t, \dots, Y_{t-p+1}\}$ up to horizon $1+q-p$. Finally, it generalizes the VAR(m) and thus produces unbiased estimates if the DGP is indeed a VAR(m).

These are just a few examples. The full range of what the VAR sequence can achieve has yet to be discovered.

5.3. *Trends and Permanent Effects*

One advantage of generalizing both LPs and VARs as VAR sequences is that it can be designed to match the characteristics of one model while leveraging tools from another. This is particularly beneficial for local projections, given that the VAR toolbox is much larger compared to LPs, which have gained popularity only recently. Below, I explore two VAR tools that can be integrated into VAR sequences, with additional details provided in Appendix A.

Firstly, a common extension of VARs is the inclusion of trends through cointegrating relationships, known as vector error correction models (VECMs), as developed by Granger (1981), Engle and Granger (1987), and Johansen (1991). In Appendix A, I define the VEC sequence as a series of VECMs, analogous to the VAR sequence. Theorem A.1 shows that a VEC sequence of increasing order replicates a local projection that incorporates the same error correction terms, thereby extending the capability of cointegration into the realm of local projections.

Secondly, while stable VARs show diminishing effects of Y_t on Y_{t+h} as h increases, VECMs or cumulated VARs can exhibit non-zero effects at the infinite horizon. These permanent effects are instrumental in identifying IRFs through long-run restrictions, as proposed by Blanchard and Quah (1989). Theorem A.2 outlines how to compute such permanent effects using a VEC sequence (or cumulated VAR sequence), under the condition that the lag order eventually stabilizes, a characteristic of VARs but not LPs. Notably, the LP-VAR($p, p+H-1$) model from Definition 2 preserves the same LP-IRFs up to horizon H, but unlike an LP, Theorem A.2 is applicable because its lag order stabilizes after horizon H. This allows for the identification of IRFs via long-run restrictions within a local

projection framework, similar to how Blanchard and Quah (1989) applied it to VARs. Consequently, the VAR sequence extends this important tool for identifying IRFs via long-run restrictions from VARs to LPs.

6. CONCLUSION

This paper establishes a formal equivalence between local projections (LPs) and vector autoregressions (VARs), demonstrating that LPs can be represented as combinations of equal and higher-order VARs, while VARs are composed of equal and lower-order LPs (Theorem 1). Consequently, LPs are more general than VARs of the same order, explaining the bias-variance tradeoff observed in the literature (Corollary 1). However, LPs, despite their generality, cannot eliminate VAR bias and are therefore less efficient than VARs under the assumption of unbiasedness (Corollary 2).

Three key applications of Theorem 1 are explored. First, it improves model selection by enabling an "apples to apples" comparison by converting LP parameters to VAR parameters before comparing LPs to VARs. By defining model size as the number of VAR parameters, it becomes clear that while there is a bias-variance tradeoff between smaller and larger models, there is no evidence of such a tradeoff between LPs and VARs once size is controlled for. Second, Theorem 1 improves LP estimation by utilizing the underlying VARs. Third, the VAR sequence (VAR-S) is introduced as a new empirical model that generalizes LPs and VARs, allowing LPs to utilize VAR tools such as modeling cointegrating relationships or identifying impulse response functions (IRFs) via long-run restrictions.

Each of the three applications suggests several promising areas for future research. For the first application, while the literature has identified scenarios where LPs outperform VARs and vice versa, these findings may stem from comparing LPs to VARs of equal order, making the former a generalization of the latter. Therefore, a comparison of LPs and VARs of equal size rather than order is needed to verify or reject these claims and uncover new scenarios that favor one model over the other.

The second application shows that efficiency can be improved by decomposing LPs into one-step VAR predictions and then estimating those predictions rather than LPs directly. Future research could explore whether such decomposition is also possible for other, potentially non-linear, direct forecasting models, potentially leading to efficiency gains for those models as well.

The third application suggests three areas for further exploration. First, while LPs and VARs are VAR sequences with increasing and constant lag orders respectively, alternative lag orders should be explored to identify the best VAR sequences for different situations. Second, while this paper demonstrates how to use the VAR sequence to transfer some of the VAR tools to LPs, future research could explore transferring other tools, such as forecast error variance decomposition (FEVD) and inference techniques. Third, by reformulating research on estimation, inference, or identification techniques in terms of the more general VAR sequence, it is possible to advance LP and VAR methods simultaneously, eliminating the need to explore these topics separately for each model.

REFERENCES

ADÄMMER, PHILIPP (2019): "lpirfs: An R Package to Estimate Impulse Response Functions by Local Projections," *The R Journal*, 11 (2), 421–438. [10]

ANG, ANDREW, MONIKA PIAZZESI, AND MIN WEI (2006): "What Does the Yield Curve Tell us About GDP Growth?" *Journal of Econometrics*, 131 (1), 359–403. [6]

AUERBACH, ALAN J. AND YURIY GORODNICHENKO (2013a): "Fiscal Multipliers in Recession and Expansion," in *Fiscal Policy after the Financial Crisis*, ed. by Alberto Alesina and Francesco Giavazzi, Chicago: University of Chicago Press, chap. 2. [2, 9, 10]

AUERBACH, ALAN J AND YURIY GORODNICHENKO (2013b): "Output Spillovers from Fiscal Policy," *The American Economic Review*, 103 (3), 141–146. [9, 10]

——— (2016): "Effects of Fiscal Shocks in a Globalized World," *IMF Economic Review*, 64 (1), 177–215. [9, 10]

AUSTIN, BENJAMIN ALFRED (2020): "Essays on Labor Economics and Econometrics; Chapter 2: "The trade-off between LP-IV and SVAR-IV estimation"," Ph.D. thesis, Harvard University. [2, 6]

BAEK, CHAEWON AND BYOUNGCHAN LEE (2022): "A Guide to Autoregressive Distributed Lag Models for Impulse Response Estimations," *Oxford Bulletin of Economics and Statistics*, 84 (5), 1101–1122. [11]

BANERJEE, ANINDYA, JUAN DOLADO, JOHN GALBRAITH, AND DAVID F. HENDRY (1993): *Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press. [11]

BARNICHON, REGIS AND CHRISTIAN BROWNLEES (2019): "Impulse Response Estimation by Smooth Local Projections," *The Review of Economics and Statistics*, 101 (3), 522–530. [6]

BERNANKE, BEN S (1986): "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25 (Supplement C), 49–99. [1, 13]

BHANSALI, RAJENDRA J. (1996): "Asymptotically Efficient Autoregressive Model Selection for Multistep Prediction," *Annals of the Institute of Statistical Mathematics*, 48 (3), 577–602. [9, 23]

——— (1997): "Direct Autoregressive Predictors for Multistep Prediction: Order Selection and Performance Relative to the Plug In Predictors," *Statistica Sinica*, 7 (2), 425–449. [9]

——— (1999): "Parameter Estimation and Model Selection for Multistep Prediction of a Time Series: A Review," in *Asymptotics, Non-Parametrics and Time Series*, ed. by Subir Ghosh, New York: Marcel Dekker, 201–225. [5, 23]

BLANCHARD, OLIVIER JEAN AND DANNY QUAH (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *The American Economic Review*, 79 (4), 655–673. [4, 7, 13, 34, 38, 39, Supp. 1, Supp. 7, Supp. 10, Supp. 11]

BLANCHARD, OLIVIER JEAN AND MARK W. WATSON (1986): *Are Business Cycles All Alike?*, Chicago: University of Chicago Press, 123–180. [1, 13]

BRUGNOLINI, LUCA (2018): "About Local Projection Impulse Response Function Reliability," CEIS Research Paper 440, Tor Vergata University, CEIS. [2, 6]

BRUNS, MARTIN AND HELMUT LÜTKEPOHL (2022): "Comparison of Local Projection Estimators for Proxy Vector Autoregressions," *Journal of Economic Dynamics and Control*, 134, 104277. [2, 6]

CHANG, PAO-LI AND SHINICHI SAKATA (2007): "Estimation of Impulse Response Functions Using Long Autoregression," *The Econometrics Journal*, 10 (2), 453–469. [7, 9, 10, 15]

CHEVILLON, GUILLAUME (2007): "Direct Multi-Step Estimation and Forecasting," *Journal of Economic Surveys*, 21 (4), 746–785. [2, 5, 24]

CHEVILLON, GUILLAUME AND DAVID F. HENDRY (2005): "Non-Parametric Direct Multi-Step Estimation for Forecasting Economic Processes," *International Journal of Forecasting*, 21 (2), 201–218. [6]

CHOI, CHI-YOUNG AND ALEXANDER CHUDIK (2019): "Estimating Impulse Response Functions When the Shock Series Is Observed," *Economics Letters*, 180 (C), 71–75. [2, 6]

COX, DAVID R. (1961): "Prediction by Exponentially Weighted Moving Averages and Related Methods," *Journal of the Royal Statistical Society: Series B*, 23 (2), 414–422. [5, 8]

ENGLE, ROBERT F. AND CLIVE W. J. GRANGER (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55 (2), 251. [4, 7, 34, 38, Supp. 1]

FAVARA, GIOVANNI AND JEAN IMBS (2015): "Credit Supply and the Price of Housing," *The American Economic Review*, 105 (3), 958–992. [10]

FINDLEY, DAVID F. (1983): "On the Use of Multiple Models for Multi-period Forecasting," in *Proceedings of Business and Economic Statistics*, American Statistical Association, 528–531. [6]

——— (1985): "Model Selection for Multi-Step-Ahead Forecasting," *IFAC Proceedings Volumes*, 18 (5), 1039–1044. [6]

FRISCH, RAGNAR AND FREDERICK V. WAUGH (1933): "Partial Time Regressions as Compared with Individual Trends," *Econometrica*, 1 (4), 387–401. [Supp. 15]

FUHRER, JEFFREY C. AND GEORGE R. MOORE (1995a): "Inflation Persistence," *The Quarterly Journal of Economics*, 110 (1), 127–159. [Supp. 12]

——— (1995b): "Monetary Policy Trade-Offs and the Correlation Between Nominal Interest Rates and Real Output," *The American Economic Review*, 85 (1), 219–239. [Supp. 12]

GALÍ, JORDI (1992): "How Well Does the IS-LM Model Fit Postwar U.S. Data?" *The Quarterly Journal of Economics*, 107 (2), 709–738. [Supp. 12]

GERSCH, WILL AND GENSHIRO KITAGAWA (1983): "The Prediction of Time Series With Trends and Seasonalities," *Journal of Business and Economic Statistics*, 1 (3), 253–264. [5]

GORODNICHENKO, YURIY AND BYOUNGCHAN LEE (2020): "Forecast Error Variance Decompositions with Local Projections," *Journal of Business and Economic Statistics*, 38 (4), 921–933. [6, 7]

GRANGER, CLIVE W. J. (1981): "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16 (1), 121–130. [38]

HAMILTON, JAMES D (1994): *Time series analysis*, Princeton: Princeton University Press. [11]

——— (2011): "Nonlinearities and the Macroeconomic Effects of Oil Prices," *Macroeconomic Dynamics*, 15 (3), 364–378. [10]

HANSEN, PETER REINHARD (2005): "Granger's Representation Theorem: A Closed-Form Expression for I(1) Processes," *The Econometrics Journal*, 8 (1), 23–38. [Supp. 9]

HENDRY, DAVID F. (1995): *Dynamic Econometrics*, Oxford University Press. [11]

ING, CHING-KANG (2003): "Multistep Prediction In Autoregressive Processes," *Econometric Theory*, 19 (2), 254–279. [9]

INOUE, ATSUSHI AND LUTZ KILIAN (2002): "Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in VAR ∞ Models," *International Economic Review*, 43 (2), 309–331. [24]

JOHANSEN, SØREN (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59 (6), 1551–1580. [7, 38, Supp. 1]

——— (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press. [11]

JOHNSTON, H. N. (1974): "A Note on the Estimation and Prediction Inefficiency of "Dynamic" Estimators," *International Economic Review*, 15 (1), 251–255. [24]

JORDÀ, ÒSCAR (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95 (1), 161–182. [1, 2, 6, 8, 9, 20, 33, Supp. 12, Supp. 13]

——— (2009): "Simultaneous Confidence Regions for Impulse Responses," *The Review of Economics and Statistics*, 91 (3), 629–647. [9]

JORDÀ, ÒSCAR, MORITZ SCHULARICK, AND ALAN M. TAYLOR (2013): "When Credit Bites Back," *Journal of Money, Credit and Banking*, 45 (2), 3–28. [10]

——— (2015): "Betting the House," *Journal of International Economics*, 96 (Supplement 1), S2–S18. [10]

——— (2020): "The Effects of Quasi-Random Monetary Experiments," *Journal of Monetary Economics*, 112, 381–419. [10]

JORDÀ, ÒSCAR AND ALAN M. TAYLOR (2016): "The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy," *The Economic Journal*, 126 (590), 219–255. [10]

KABAILA, PAUL V. (1981): "Estimation Based on One Step Ahead Prediction Versus Estimation Based on Multi-Step Ahead Prediction," *Stochastics*, 6 (1), 43–55. [24]

KANG, IN-BONG (2003): "Multi-Period Forecasting Using Different Models for Different Horizons: An Application to U.S. Economic Time Series Data," *International Journal of Forecasting*, 19 (3), 387–400. [6]

KILIAN, LUTZ AND YUN JUNG KIM (2011): "How Reliable Are Local Projection Estimators of Impulse Responses?" *The Review of Economics and Statistics*, 93 (4), 1460–1466. [2, 6]

KILIAN, LUTZ AND HELMUT LÜTKEPOHL (2017): *Structural Vector Autoregressive Analysis*, Structural Vector Autoregressive Analysis, Cambridge University Press. [5]

KING, ROBERT G., CHARLES I. PLOSSER, JAMES H. STOCK, AND MARK W. WATSON (1991): "Stochastic Trends and Economic Fluctuations," *The American Economic Review*, 81 (4), 819–840. [7]

LEDUC, SYLVAIN AND DANIEL J WILSON (2017): "Are State Governments Roadblocks to Federal Stimulus? Evidence on the Flypaper Effect of Highway Grants in the 2009 Recovery Act," *American Economic Journal: Economic Policy*, 9 (2), 253–292. [10]

LEWIS, RICHARD AND GREGORY C REINSEL (1985): "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16 (3), 393–411. [3, 24]

LI, DAKE, MIKKEL PLAGBORG-MØLLER, AND CHRISTIAN K. WOLF (2023): "Local Projections vs. VARs: Lessons From Thousands of DGPs," Working paper. [2, 6]

LIN, JIN-LUNG AND RUEY S. TSAY (1996): "Co-Integration Constraint and Forecasting: An Empirical Examination," *Journal of Applied Econometrics*, 11 (5), 519–538. [9]

LIU, SHU-ING (1996): "Model Selection for Multiperiod Forecasts," *Biometrika*, 83 (4), 861–873. [6]

LOVELL, MICHAEL C (1963): "Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis," *Journal of the American Statistical Association*, 58 (304), 993–1010. [Supp. 15]

MARCELLINO, MASSIMILIANO, JAMES H STOCK, AND MARK W WATSON (2006): "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series," *Journal of Econometrics*, 135 (1), 499–526. [2, 5, 6]

MCELROY, TUCKER (2015): "When are Direct Multi-step and Iterative Forecasts Identical?" *Journal of Forecasting*, 34 (4), 315–336. [2, 5]

NAKAMURA, EMI AND JÓN STEINSSON (2018): "Identification in Macroeconomics," *The Journal of Economic Perspectives*, 32 (3), 59–86. [1, 5, 10, 11]

NUSBAUM, CHARLIE (2020): "Aggregate Hours and Local Projections with Long-Run Restrictions," *SSRN Electronic Journal*, 1–31. [7, Supp. 10, Supp. 12]

OWYANG, MICHAEL T., VALERIE A. RAMEY, AND SARAH ZUBAIRY (2013): "Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data," *The American Economic Review*, 103 (3), 129–134. [10]

PESARAN, M. HASHEM, ANDREAS PICK, AND ALLAN TIMMERMANN (2011): "Variable Selection, Estimation and Inference for Multi-Period Forecasting Problems," *Journal of Econometrics*, 164 (1), 173–187. [2, 5, 6, 31]

PFAFF, BERNHARD (2008): *Analysis of Integrated and Cointegrated Time Series with R*, New York: Springer, second ed. [11]

PLAGBORG-MØLLER, MIKKEL AND CHRISTIAN K. WOLF (2021a): "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, 89 (2), 955–980. [1, 2, 5, 6, 7, 10, 11, 15, 19, 20, 21, 22, 27]

——— (2021b): "Supplement to 'Local Projections and VARs Estimate the Same Impulse Responses'," . [15]

——— (2022): "Instrumental Variable Identification of Dynamic Variance Decompositions," *Journal of Political Economy*, 130 (8), 2164–2202. [7]

RAMEY, VALERIE A. (2016): "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, ed. by John B. Taylor and Harald Uhlig, Elsevier, vol. 2, chap. 2, 71–162. [1, 5, 6, 10, 11]

RAMEY, VALERIE A AND SARAH 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. [10]

ROMER, CHRISTINA D AND DAVID H ROMER (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94 (4), 1055–1084. [11]

——— (2010): "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 100 (3), 763–801. [11]

RUDEBUSCH, GLENN AND LARS E. O. SVENSSON (1999): "Policy Rules for Inflation Targeting," in *Monetary Policy Rules*, National Bureau of Economic Research, Inc, NBER Chapters, 203–262. [Supp. 12]

SCHORFHEIDE, FRANK (2005): "VAR Forecasting Under Misspecification," *Journal of Econometrics*, 128 (1), 99–136. [2, 5, 6]

SHAPIRO, MATTHEW AND MARK WATSON (1988): "Sources of Business Cycles Fluctuations," in *NBER Macroeconomics Annual 1988, Volume 3*, National Bureau of Economic Research, Inc, NBER Chapters, 111–156. [7]

SIMS, CHRISTOPHER A. (1980): "Macroeconomics and Reality," *Econometrica*, 48 (1), 1–48. [1, 11]

SIMS, CHRISTOPHER A. AND TAO ZHA (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39 (4), 949–968. [13]

STOCK, JAMES H. AND MARK W. WATSON (2018): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *The Economic Journal*, 128 (610), 917–948. [5, 10]

SVETUNKOV, IVAN, NIKOLAOS KOURENTZES, AND REBECCA KILLICK (2023): "Multi-Step Estimators and Shrinkage Effect in Time Series Models," *Computational Statistics*. [5]

TAYLOR, JOHN B. (1999): *Monetary Policy Rules*, National Bureau of Economic Research, University of Chicago Press. [Supp. 12]

TENREYRO, SILVANA AND GREGORY THWAITES (2016): "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions," *American Economic Journal: Macroeconomics*, 8 (4), 43–74. [10]

TIAO, GEORGE C. AND RUEY S. TSAY (1994): "Some Advances in Non-Linear and Adaptive Modelling in Time-Series," *Journal of Forecasting*, 13 (2), 109–131. [6]

TIAO, GEORGE C. AND DAMING XU (1993): "Robustness of Maximum Likelihood Estimates for Multi-Step Predictions: The Exponential Smoothing Case," *Biometrika*, 80 (3), 623–641. [5]

TSAY, RUEY S. (1993): "Comment: Adaptive Forecasting," *Journal of Business and Economic Statistics*, 11 (2), 140–142. [9]

UHLIG, HARALD (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52 (2), 381–419. [13]

WEISS, ANDREW A (1991): "Multi-Step Estimation and Forecasting in Dynamic Models," *Journal of Econometrics*, 48 (1), 135–149. [6, 8]

Supplement to "Local Projections Are VAR Predictions of Increasing Order"

Julian F. Ludwig

This online appendix provides supplementary material for the article “Local Projections Are VAR Predictions of Increasing Order.” Any references to sections, tables, figures, and equations without the prefixes “A.”, “B.”, or “C.” correspond to the main article.

APPENDIX A: VEC SEQUENCE - THE INCLUSION OF TRENDS

Incorporating trends—both deterministic and stochastic—is essential for modeling and forecasting. This section extends the framework of LPs, VARs, and VAR sequences to include these trends, building on the vector error correction model (VECM), a critical model introduced by Engle and Granger (1987) and further developed by Johansen (1991), for modeling non-stationary time series and estimating long-term effects. The main contributions are developing an error correction version of both, a local projection and the new model introduced in Section 5, called VAR sequence.

Sections A.1 and A.2 introduce deterministic and stochastic trends, respectively. Section A.3 defines the VECM, the VEC local projection, and the VEC sequence as a trend-augmented VAR, LP, and VAR sequence. Section A.4 demonstrates that the VEC sequence is a generalization and extends Theorem 1 by showing that VECLPs are VECM predictions of increasing order. Section A.5 shows how to compute permanent effects using VEC sequences and applies those insights to identifying IRFs via long-run restrictions á la Blanchard and Quah (1989), not only when estimated with VARs but also for LPs, something that has not previously been possible.

A.1. *Deterministic Trend*

I consider two types of trends in the N-dimensional stochastic process $\{Y_t\}$: deterministic and stochastic trends. The deterministic trend is a polynomial of degree M in time t,

following the process:

$$\delta_t^M = D_M \delta_{t-1}^M, \qquad \delta_t^M = \begin{bmatrix} t^0 \\ \vdots \\ t^M \end{bmatrix} \qquad [D_M]_{ij} = \frac{(i-1)!}{(i-j)!(j-1)!} \tag{A.1}$$

where D_M is an invertible $(M+1)\times(M+1)$ matrix, with $[D_M]_{ij}$ representing the element in row i and column j, and ! denoting the factorial operator. Note that differencing the trend reduces the degree by one; hence, the deterministic trend of the differenced process $\{\Delta Y_t\}$ is of degree $M-1$.

If the trend in $\{Y_t\}$ is solely deterministic of degree M, the VAR in (6) can account for this by replacing the N-dimensional vector of intercepts $\gamma(p)$ with $\gamma_\delta(p)\delta_t^M$, where δ_t^M is the deterministic variable defined in (A.1) and $\gamma_\delta(p)$ is an $N \times (M+1)$ matrix of coefficients to be estimated. This implies that the intercepts $\pi^{(h)}(p)$ of the prediction version of the VAR in (8) are replaced with $\pi_\delta^{(h)}(p)\delta_t^M$, given the following $N \times (M+1)$ matrix of coefficients $\pi_\delta^{(h)}(p)$:

$$\pi_\delta^{(h)}(p) = \gamma_\delta(p) D_M^h + \sum_{k=1}^{\min\{h-1,p\}} \Gamma_k(p)\pi_\delta^{(h-k)}(p) \qquad h = 1,\dots,H \tag{A.2}$$

where $\Gamma_k(p)$ are the $N \times N$ slope coefficients in (6) after replacing $\gamma(p)$ with $\gamma_\delta(p)\delta_t$. Similarly, replacing $b^{(h)}(p)$ with $b_\delta^{(h)}(p)\delta_t^M$ in the local projection in (1) accounts for the deterministic trend in that model. It is straightforward to show that Theorem 1 still holds when such a trend is included.

A.2. *Stochastic Trend and Cointegration*

Suppose the N-dimensional stochastic process $\{Y_t\}$ contains $N-R$ stochastic trends, meaning both $\Delta Y_t = Y_t - Y_{t-1}$ as well as R linear combinations of Y_t are (trend-)stationary. Those linear combinations are called cointegrating relationships, denoted as $\beta' Y_t$, where β is an $N \times R$ matrix with rank $R \in \{0,\dots,N\}$. The remaining combinations become (trend-)stationary only after differencing the time series.

If there are as many cointegrating relationships as variables ($R = N$), then all linear combinations are trend-stationary, including $\beta = I_N$, and a standard VAR, LP, or VAR-S can be used with the inclusion of a deterministic trend of degree M. If none of the linear combinations of Y_t are trend-stationary, the models can be estimated for the differenced series ΔY_t with the inclusion of a deterministic trend of degree $M - 1$.

More generally, regardless of the number of cointegrating relationships, the models can be estimated using the differenced series with the inclusion of a deterministic trend of degree $M - 1$ as well as the following cointegrating relationships that are trend-stationary of degree $M - 1$:

$$Z_t = \begin{bmatrix} \beta' & \rho \end{bmatrix} \begin{bmatrix} Y_t \\ t^M \end{bmatrix} \tag{A.3}$$

where β is an $N \times R$ matrix and ρ is an R-dimensional vector of coefficients capturing the cointegrating relationships. Here, $R = 0$ results in $Z_t = 0$ and $R = N$ results in $Z_t = \begin{bmatrix} Y_t' & t^M \end{bmatrix}'$. The inclusion of t^M in the cointegrating relationships means that the stochastic process $\{Y_t\}$ may include as many independent deterministic trends of degree M as variables. The models that incorporate such cointegrating relationships are referred to as models with a vector error correction (VEC) term, as introduced in the next section.

A.3. *Definition of VECM, VECLP, and VEC Sequence*

Vector Error Correction Model (VECM)

The vector error correction model of order p, denoted as VECM(p), extends the VAR in (6) to account for the N deterministic and $N - R$ stochastic trends of Section A.1 and A.2 as follows:

$$\Delta Y_t = \phi_\delta(p)\delta_t^{M-1} + \alpha(p)Z_{t-p} + \Phi_1(p)\Delta Y_{t-1} + \cdots + \Phi_{p-1}(p)\Delta Y_{t-p+1} + \eta_t(p) \tag{A.4}$$

where δ_t^{M-1} is an M-dimensional vector of deterministic trends as per equation (A.1), Z_{t-p} are the cointegration relationships defined in (A.3), $\phi_\delta(p)$ is an $N \times M$ vector of constants, and $\Phi_l(p)$ is an $N \times N$ matrix of coefficients for each lag l. By definition of a regression, the VECM residuals $\eta_t(p)$ have a mean of zero and are not correlated with regressors. Replacing Z_{t-p} with Z_{t-1} gives the same predictions and is a common practice

in these models. Note that the processes of differencing and cointegration ensure that all components of the regression are trend-stationary, resulting in consistent estimates.

The VECM(p) can be converted to a VAR(p) with a deterministic trend of degree M by using the definitions of ΔY_t and Z_{t-p}. After this conversion, the VECM(p) forecasts and IRFs are obtained by iterating the VAR(p) forward as in Section 2.1, with a trend component as per equation (A.2). Note that these VECM predictions generalize both the VAR($p-1$) predictions with differenced variables when $R = 0$ and the standard VAR(p) predictions with level variables when $R = N$ (see Theorem A.1).

VEC Local Projection (VECLP)

A deterministic trend and a vector error correction (VEC) term can also be included in the local projection of (1) with differenced time series, similar to the VECM in (A.4). However, in the spirit of local projections, which estimate coefficients directly, the following model, denoted as VECLP(p), is the most direct approach to estimating LP coefficients while keeping all variables trend-stationary:

$$\begin{aligned} Y_{t+h} - Y_{t-p+1} = o_\delta^{(h)}(p)\delta_t^{M-1} + a^{(h)}(p)Z_{t-p+1} + C_1^{(h)}(p)(Y_t - Y_{t-p+1}) \\ + \cdots + C_{p-1}^{(h)}(p)(Y_{t-p+2} - Y_{t-p+1}) + u_{t+h}^{(h)}(p) \qquad h = 1, 2, \ldots \end{aligned} \tag{A.5}$$

where $o_\delta^{(h)}(p)$ is an $N \times M$ vector of constants, and $C_l^{(h)}(p)$ is an $N \times N$ matrix of coefficients for each lag l. By definition of a regression, the VECLP residuals $u_{t+h}^{(h)}(p)$ have a mean of zero and are not correlated with regressors.

The estimated local projection in (A.5) converts to the following prediction model:

$$Y_{t+h} = c_\delta^{(h)}(p)\delta_t^M + C_1^{(h)}(p)Y_t + \cdots + C_p^{(h)}(p)Y_{t-p+1} + u_{t+h}^{(h)}(p) \qquad h = 1, 2, \ldots \tag{A.6}$$

where $c_\delta^{(h)}(p) = \left[o_\delta^{(h)}(p)\, 0\right] + \left[0\, a^{(h)}(p)\rho\right] D_M^{1-p}$, and $C_p^{(h)}(p) = I_N + a^{(h)}(p)\beta' - \sum_{l=1}^{p-1} C_l^{(h)}(p)$. The VECLP($p$) forecasts and IRFs are then obtained as in (2) and (3).

Note that all coefficients of the prediction model except $c_\delta^{(h)}(p)$ and $C_p^{(h)}(p)$ are estimated directly, including the coefficients used for IRF estimation $C_1^{(h)}(p)$. Alternatively, one could replace the dependent variables $Y_{t+h} - Y_{t-p+1}$ with ΔY_{t+h} or $Y_{t+h} - Y_t$, which results in the same prediction model in (A.6), due to the linear properties of covariance. For

example, it is common to use $Y_{t+h} - Y_t$ as the dependent variable to directly estimate the IRF coefficients $C_1^{(h)}(p)$; however, with this specification, the coefficients at lag $l > 1$ need to be derived indirectly.

Theorem A.1, introduced later in this section, shows that the VECLP model in (A.6) is identical to cumulating VECM predictions of increasing order, thus extending Theorem 1 to time series with stochastic trends.

VEC Sequence (VEC-S)

Analogous to the VAR sequence (VAR-S) for stationary processes, the VEC sequence (VEC-S) is a new method that estimates impulse response functions (IRFs) and forecasts for processes with stochastic trends, based on a series of vector error correction models (VECMs). The VEC-S(s), defined by a particular *lag sequence* $s = \{s_1, \dots, s_H\}$ or denoted as VEC-S($s_1, \dots, s_H$), operates by estimating a VECM at each lag order $s_h \in s$. These individual VECMs are then combined to produce IRFs and forecasts through a four-step procedure, similar to the procedure used to compute the VAR-S in Section 5.1.

In the first step, the VECMs of order s_j, for $j = 1, \dots, h$, are listed in order, while shifting the time index incrementally:

$$\begin{aligned} \Delta Y_{t+j} = \phi_\delta(s_j)\delta_{t+j}^{M-1} + \alpha(s_j)Z_{t+j-s_j} \\ + \Phi_1(s_j)\Delta Y_{t+j-1} + \cdots + \Phi_{s_j-1}(s_j)\Delta Y_{t+j-(s_j-1)} + \eta_{t+j}(s_j) \qquad j = 1, \dots, h \end{aligned} \tag{A.7}$$

where $\phi_\delta(s_j)$, $\alpha(s_j)$, $\Phi_l(s_j)$, and $\eta_t(s_j)$ are VECM(s_j) coefficients and residuals (A.4).

In the second step, the VECMs are combined to replace future regressors with current and past variables:

$$\begin{aligned} \Delta Y_{t+h} = \omega_\delta^{(h)}(s)\delta_t^{M-1} + \kappa^{(h)}(s)Z_{t-\lambda_s+1} \qquad & h = 1, 2, \dots \\ + \Omega_1^{(h)}(s)\Delta Y_t + \cdots + \Omega_{\lambda_s-1}^{(h)}(s)\Delta Y_{t-\lambda_s+2} + \chi_{t+h}^{(h)}(s) \qquad & \lambda_s = \max_h\{s_h + 1 - h\} \end{aligned} \tag{A.8}$$

where the coefficients and forecast errors are functions of the VECMs in (A.7). To simplify notation, define $\Phi_k(s_h) = \alpha(s_h)\beta'$ for $k \geq s_h$ and $\varphi_\delta(s_h) = \phi_\delta(s_h)D_{M-1}^h + \left[0\ \alpha(s_h)\rho\right]\left(D_M^{h-s_h}\right.$

$-D_M^{1-\lambda_s}\Big)\left[I_M\,0\right]'$ so that the components of (A.8) are as follows:

$$\Omega_l^{(h)}(s)=\sum_{k=1}^{h+\lambda_s-2}\Phi_k(s_h)\Omega_l^{(h-k)}(s) \qquad \kappa^{(h)}(s)=\alpha(s_h)+\sum_{k=1}^{h-1}\Phi_k(s_h)\kappa^{(h-k)}(s)$$

$$\omega_\delta^{(h)}(s)=\varphi_\delta(s_h)+\sum_{k=1}^{h-1}\Phi_k(s_h)\omega_\delta^{(h-k)}(s) \qquad \chi_{t+h}^{(h)}(s)=\eta_{t+h}(s_h)+\sum_{k=1}^{h-1}\Phi_k(s_h)\chi_{t+h-k}^{(h-k)}(s)$$

$$l=1,\dots,\lambda_s-1 \qquad\qquad h=1,\dots,H \tag{A.9}$$

with normalization $\Omega_l^{(1-k)}(s)=I_N$ for all $l=k\geq 1$, and $\Omega_l^{(1-k)}(s)=0$ otherwise.

In the third step, the prediction model for ΔY_{t+h} in (A.8) is cumulated across horizon h and Y_t is added on both sides to derive the following expression for the level variables:

$$Y_{t+h}=f_\delta^{(h)}(s)\delta_t^M+F_1^{(h)}(s)Y_t+\cdots+F_{\lambda_s}^{(h)}(s)Y_{t-\lambda_s+1}+w_{t+h}^{(h)}(s) \qquad h=1,2,\dots \tag{A.10}$$

where coefficients and forecast errors are the following functions of the model in (A.8):

$$F_l^{(h)}(s)=\sum_{k=1}^{h}\begin{cases} I_N+\Omega_1^{(k)}(s) & \text{if } l=1\\ \Omega_l^{(k)}(s)-\Omega_{l-1}^{(k)}(s) & \text{if } 1<l<\lambda_s\\ \kappa^{(k)}(s)\beta'-\Omega_{\lambda_s-1}^{(k)}(s) & \text{if } l=\lambda_s\end{cases} \qquad w_{t+h}^{(h)}(s)=\sum_{k=1}^{h}\chi_{t+k}^{(k)}(s)$$

$$f_\delta^{(h)}(s)=\sum_{k=1}^{h}\left(\left[\omega_\delta^{(k)}(s)\,0\right]+\left[0\,\kappa^{(k)}(s)\rho\right]D_M^{1-\lambda_s}\right) \qquad h=1,2,\dots \tag{A.11}$$

with normalizations $F_l^{(1-k)}(s)=I_N$, for all $l\geq k\geq 1$, and $F_l^{(1-k)}(s)=0$ otherwise.

The final step is to obtain multistep forecasts by removing the forecast errors in (A.10):

$$Y_{t+h|t}=f_\delta^{(h)}(s)\delta_t^M+F_1^{(h)}(s)Y_t+\cdots+F_{\lambda_s}^{(h)}(s)Y_{t-\lambda_s+1} \qquad h=1,\dots,H \tag{A.12}$$

which then implies the following impulse response functions (IRFs):

$$IRF(h,d)=F_1^{(h)}(s)d \qquad h=0,\dots,H \tag{A.13}$$

with normalization $F_1^{(0)}(s) = I_N$, where h is the IRF horizon, d is the impact of the shock, and $s = \{s_1, \ldots, s_H\}$ is the lag sequence. Note that if there are no cointegrating relationships, i.e., $R = 0$ and $Z_t = 0$, then the IRFs are identical to the cumulated IRFs of a VAR-S(r) for the differenced time series, where $r_h = s_h - 1$ for all h (see Theorem A.1). Since the VEC sequence accommodates stochastic trends, the IRFs may converge to non-zero values as the horizon goes to infinity. This feature is commonly exploited in identifying the impact vector d using long-run restrictions, as shown by Blanchard and Quah (1989) and explored in Section A.5.

A.4. *Theorem A.1: VECM, VECLP, and VAR-S Are VEC Sequences*

The following theorem shows that the VEC sequence is a generalization:

THEOREM A.1—VECM, VECLP, and VAR-S Are VEC Sequences: *Consider a VAR-S(s), VECM(q), VECLP(p), and a VEC-S(s) as per (26), (A.4), (A.6), and (A.10), with $R \in \{0, \ldots, N\}$ representing the number of cointegrating relationships defined in (A.3). Then the following holds true in both population and sample:*

(a) A VEC-S($q, q, q, \ldots$) produces the same IRFs and forecasts as a VECM(q),

(b) a VEC-S($p, p+1, p+2 \ldots$) produces the same IRFs and forecasts as a VECLP(p),

(c) a VEC-S(s) with $R = N$ produces the same IRFs and forecasts as a VAR-S(s), and

(d) a VEC-S(s) with $R = 0$ produces the same IRFs and forecasts as those obtained by first estimating a VAR-S(r) for differenced time series and then cumulating the coefficients to obtain level predictions, where $r_h = s_h - 1$ for all h.

PROOF: Statement (a) follows directly from the definition of a VEC-S($q, q, q, \ldots$) as a sequence of VECM(q) predictions. The proof of (b) parallels the proof of Theorem 1 but using VECMs instead of VARs and VECLPs instead of LPs. To prove (c), note that $R = N$ implies $Z_t = \left[Y_t' \; t^M\right]'$, so the regressors of a VECM(s_h) are an invertible linear transformation of the regressors of the VAR(s_h). Therefore, they estimate the same predictions. This holds even though the dependent variables of the VECM(s_h) are differenced ΔY_t. That's because the subtracted variables Y_{t-1} are controlled for and the linear properties of covariance apply, i.e., $\mathrm{Cov}(Y_t, Y_{t-1}) = \mathrm{Cov}(\Delta Y_t, Y_{t-1}) + \mathrm{Var}(Y_{t-1})$. Statement (d) follows

from the fact that $R = 0$ implies $Z_t = 0$, which turns the VECM(s_h) into a VAR($s_h - 1$) for differenced variables. *Q.E.D.*

Statements (c) and (d) show that VEC sequences with no meaningful cointegrating relationships (i.e., when $R = N$ or $R = 0$) reduce to VAR sequences. Meanwhile, statements (a) and (b) show that VECLPs and VECMs are both instances of VEC sequences, with the lag sequence increasing linearly for VECLPs and remaining constant for VECMs. However, as with a VAR-S(s), the lag sequence of a VEC-S(s) does not need to adhere to either of these two patterns, and it is possible that superior patterns exist that have yet to be discovered. Examples include the VEC versions of the VAR sequences outlined in Definition 2, such as the VEC:LP-VAR(p, q), which corresponds to a VEC-S($p, p+1, \ldots, q, q, \ldots$).

Note that Theorem A.1 applies as long as the models can be estimated. Therefore, the VECM assumptions that ΔY_t and Z_t are trend-stationary and that the deterministic trend in $\{Y_t\}$ is a polynomial of degree M do not need to hold for Theorem A.1 to be valid. This is because Theorem A.1 compares the predictions of different empirical models rather than comparing the predictions to the truth.

A.5. *Theorem A.2: Identification of Permanent Effects*

As the process contains $N - R$ stochastic trends, the effects of Y_{t-l} on Y_{t+h} may be non-zero as the horizon goes to infinity. The following theorem derives those permanent effects for a VEC sequence under the condition that the lag sequence eventually stabilizes at a constant value, i.e., $s_h = q$ for $h > H$:

THEOREM A.2—Identification of Permanent Effects: *Let* $A(q)[z] = I_N(1-z) - \alpha(q)\beta' z^q - \sum_{l=1}^{q-1} \Phi_l(q)\left(z^l - z^{l+1}\right)$ *be the lag polynomial of a stable VECM(q) in (A.4) with the lag operator replaced by* $z \in \mathbb{C}$*, and let* $R \in \{0, \ldots, N\}$ *represent the number of cointegrating relationships defined in (A.3). Stability of VECM(q) implies that (i) there are* $N - R$ *unit roots in the characteristic polynomial* $\det(A(p)[z])$*, with the remaining* R *roots lying outside the unit circle, and (ii) the matrix* $\alpha(p)\beta'$ *has rank* R*, with* $\alpha(p)$ *and* β *being* $N \times R$ *matrices of full column rank* R*. Then the following is true:*

(a) The VECM(q) with $R < N$ produces the following effects of Y_t on Y_{t+h} as $h \to \infty$:

$$\Xi(q) = \beta_\perp \left[\alpha(q)'_\perp \left(I_N - \sum_{j=1}^{q-1} \Phi_j(q) \right) \beta_\perp \right]^{-1} \alpha(q)'_\perp$$

whereas $R = N$ implies $\Xi(q) = 0$. Here, $\alpha(p)_\perp$ and $\beta_\perp$ are the orthogonal complements of $\alpha(p)$ and β for $R > 0$, and $\beta_\perp = \alpha(p)_\perp = I_N$ for $R = 0$.

(b) The VEC-S($s_1, \ldots, s_H,\ q, q, \ldots$) defined in (A.10) produces the following effects of Y_{t+1-l} on Y_{t+h} as $h \to \infty$:

$$F_l^{(\infty)}(s) = \Xi(q) \left(F_l^{(H)}(s) - \sum_{k=1}^{q-1} \Phi_k(q) F_l^{(H-k)}(s) \right) \qquad l = 1, \ldots, \lambda_s$$

$$s = \{s_1, \ldots, s_H, q, q, \ldots\}$$

Here, $F_l^{(h)}(s)$ are the coefficients of the VEC-S(s) in (A.10), where the normalizations are given by $F_l^{(1-k)}(s) = I_N$ for all $l \geq k \geq 1$, and $F_l^{(1-k)}(s) = 0$ otherwise. The matrix $\Xi(q)$ and, consequently, $F_l^{(\infty)}(s)$ have rank $N - R$.

PROOF: Note that (a) is a special case of (b) when $H = 0$ and $l = 1$, so it suffices to prove (b). According to Theorem 1 in Hansen (2005), assumptions (i) and (ii), combined with the condition that all elements of the deterministic trend satisfy $\delta_{it}^M < a + |t|^b$ for some constants $a, b \in \mathbb{R}$, ensure that Granger's representation theorem applies to the stable VECM(q). This theorem expresses the predictions Y_{t+h} as a function of a deterministic trend, initial values Y_{t-l} for $l \geq 0$, and two terms depending on VECM residuals—one representing permanent effects and the other representing temporary effects. By definition, a VEC-S($s_1, \ldots, s_H,\ q, q, \ldots$) allows predictions to be decomposed into h-step predictions for $h \leq H$ and VECM(q) predictions that iterate those predictions forward. Thus, the expression in Theorem A.2 is derived by converting those h-step predictions into corresponding predictions of the initial values and residuals, applying Granger's representation theorem to those values, and rewriting the predictions of the initial values and residuals back as functions of the VEC-S(s) regressors. *Q.E.D.*

Theorem A.2 reveals that when the lag sequence stabilizes at a constant after a finite horizon, it enables the estimation of permanent effects, making VEC sequences with this property particularly attractive. A common application of estimating permanent effects is the identification of impulse response functions using long-run restrictions, as illustrated in the example at the end of this section.

Theorem A.1 reveals that the lag sequence of a local projection increases continually; therefore, Theorem A.2 does not apply to VECLPs, and their permanent effects cannot be estimated. This is a significant drawback of local projections compared to VARs. Previous literature has attempted to address this issue by setting $F_l^{(\infty)}(s) = F_l^{(H)}(s)$ for some finite H (Nusbaum, 2020, e.g., see). However, Theorem A.1 and Theorem A.2 offer a more elegant solution: a VEC sequence can be used to match the predictions of a VECLP(p) in (A.6) for all horizons $h \leq H$ but then stabilize the lag sequence beyond horizon H, allowing for the estimation of permanent effects. This approach corresponds to the VEC:LP-VAR($p, p + H - 1$) model of Definition 2. The following example demonstrates how this model can be used to identify IRFs via long-run restrictions.

Example

Blanchard and Quah (1989) distinguish between supply and demand shocks using long-run restrictions. They assume that only supply shocks have a permanent effect on output, while demand shocks only affect output in the short run. To estimate these effects, they use a VAR in differences, which corresponds to a VEC sequence with no cointegrating relationships (i.e., $R = 0$). To identify the two shocks, they include two variables in the system: output, measured as the log of GNP, and the unemployment rate. Specifically, their variables are defined as $\Delta Y_t = \left[\Delta \ln(\text{GNP}_t) \ \text{Unemployment}_t\right]'$.[15]

The IRF impulse vectors d_1 for supply shocks and d_2 for demand shocks are computed by decomposing the residual covariance matrix of the VECM(p) as follows:

$$d_j = \mathbf{A}e_j, \qquad \mathbf{A} = F_1^{(\infty)}(s)^{-1}\text{Chol}\Big(F_1^{(\infty)}(s)\text{Cov}\big(\eta_t(p)\big)F_1^{(\infty)}(s)'\Big) \qquad j = 1, 2 \tag{A.14}$$

[15] To compute IRFs for unemployment rather than its cumulative version, use $\left[F_1^{(h)}(s) - F_1^{(h-1)}(s)\right] d_j$. Alternatively, one could estimate a VEC sequence with $R = 1$ and $Y_t = \left[\ln(\text{GNP}_t) \ \text{Unemployment}_t\right]'$ where $\beta = \left[0 \ 1\right]'$ in (A.3), ensuring that unemployment does not have a stochastic trend.

where e_j is an N-dimensional vector with zeros except for a one in the jth position, extracting the jth column. $\mathrm{Cov}(\cdot)$ computes the covariance matrix, and $\mathrm{Chol}(\cdot)$ performs Cholesky decomposition to produce a lower-triangular matrix. $F_1^{(\infty)}(s)$ represents the effect of Y_t on Y_{t+h} as $h \to \infty$, as defined in Theorem A.2, and is invertible when $R = 0$.

Blanchard and Quah (1989) use the same VAR to estimate both the impulse vectors and the IRFs. This corresponds to a VEC-S$(p, p, \dots)$ with no cointegrating relationships. Thus, $F_1^{(\infty)}(s)$ becomes a special case of Theorem A.2 for $H = 0$, $l = 1$, and $R = 0$:

$$F_1^{(\infty)}(s) = \left(I_N - \sum_{j=1}^{q-1} \Phi_j(p) \right)^{-1} \qquad s = \{p, p, \dots\} \tag{A.15}$$

where $\Phi_j(p)$ are the VECM coefficients from (A.4), which are identical to the VAR$(p-1)$ coefficients $\Gamma_j(p-1)$ from (6) estimated for the differenced process $\{\Delta Y_t\}$.

How can this method be applied to local projections? The IRF horizon of a local projection is bounded as the horizon cannot exceed the sample size. However, Theorem A.1 shows that local projections are VECM predictions of increasing order, and it is thus reasonable to simply stop the sequence from increasing beyond a certain horizon for estimating such permanent effects. This corresponds to a VEC:LP-VAR$(p, p+H-1)$ with $R = 0$, yielding the following permanent effects according to Theorem A.2:

$$F_1^{(\infty)}(s) = \left(I_N - \sum_{j=1}^{q-1} \Phi_j(q) \right)^{-1} \left(F_1^{(H)}(s) - \sum_{k=1}^{q-1} \Phi_k(q) F_1^{(H-k)}(s) \right) \tag{A.16}$$
$$s = \{p, p+1, \dots, q, q, \dots\}, \quad q = p + H - 1$$

Here, $F_1^{(h)}(s)$ are the coefficients of the VEC-S$(p, p+1, \dots, q, q, \dots)$ in (A.10) for $q = p + H - 1$, which are identical to the coefficients $C_1^{(h)}(p)$ of the VECLP(p) in (A.6) for $h \leq H$. When $R = 0$, these coefficients are also identical to the cumulated coefficients $\sum_{k=1}^{h} B_1^{(k)}(p-1)$ of the LP$(p-1)$ in (1) estimated for $\{\Delta Y_t\}$. Using this expression for $F_1^{(\infty)}(s)$ in (A.14) provides a local projection version of identifying shocks via long-run restrictions, allowing LPs to utilize a powerful tool previously only available to VARs.

To summarize, Theorem A.2 reveals that a VEC sequence that stabilizes over time is capable of making long-term predictions. Together with Theorem A.1, which shows that local projections are VEC sequences, this can be used to generate a VEC sequence that not only replicates the predictions of a local projection but also allows for the computation of permanent effects necessary for identifying shocks via long-run restrictions. This method provides a seamless progression from the estimated IRFs to the IRFs at the infinite horizon, improving upon approaches like Nusbaum (2020) that assume IRFs of differenced variables become zero beyond the maximum horizon where LP estimation is still possible. Furthermore, this section introduces an error correction version of a local projection, enabling it to utilize all the tools previously available only to vector error correction models.

APPENDIX B: EMPIRICAL EXAMPLE

An empirical example is used to achieve two objectives. The first is to demonstrate that Theorem 1 holds not just in theory but also in practice with finite data, as shown in Figure 2. The second is to demonstrate the superiority of the new LP estimation method detailed in Section 4, as presented in Figure 3. Section B.1 details the model specifications, Section B.2 explains the rolling forecasting scheme for assessing forecast performance, and Section B.3 describes the data.

B.1. *Model Specification*

This paper replicates the model specifications from Jordà's (2005) study, which were used to compare VARs with local projection impulse response functions (IRFs), using U.S. macro indicators (see Jordà, 2005, Figure 5). The replication extends the original data set to include observations from the first quarter of 1955 up to the first quarter of 2021, thereby incorporating more recent data.

Jordà's (2005) example follows the analysis of Rudebusch and Svensson (1999), who estimates monetary policy rules with a VAR. Jordà (2005) uses the three variables of a simple, new-Keynesian, closed economy model (Galí, 1992, Fuhrer and Moore, 1995a,b, Taylor, 1999, see): output gap, inflation, and interest rate, and then estimates a VAR and a local projection model, as well as a cubic projection model, which I don't include here.

The variables of the models are structured as follows:

$$Y_t = \begin{bmatrix} \text{FFR}_t \\ \text{Inflation}_t \\ \text{Output Gap}_t \end{bmatrix} \quad t = 1, \ldots, T \tag{B.1}$$

where $t = 1$ refers to 1955 $Q1$, $t = T$ refers to 2021 $Q1$, and the difference between t and $t-1$ is one quarter. The measurement of these variables are detailed in Section B.3. Following Jordà (2005), the benchmark LP(p) and VAR(p) models of expressions (1) and (6) are estimated with a lag length of four quarters, i.e. $p = 4$.

These model specifications are used to generate the forecasts displayed in Figure 2, predicting the sixteen quarters following the first quarter of 2021. Moreover, the impulse response functions (IRFs) shown in the same figure are calculated with an impulse vector of $d = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}'$, which simulates a one percentage point increase in the federal funds rate (FFR), while holding current and past variables fixed. This impulse vector identifies the same relative IRFs as Jordà's (2005) *Fed Funds shock*, which is based on an estimated impulse vector of $\hat{d} \approx \begin{bmatrix} 0.8 & 0 & 0 \end{bmatrix}'$.

B.2. *Out-Of-Sample Performance*

To evaluate the out-of-sample performance of the three LP estimation methods detailed in Section 4, a rolling forecasting scheme is employed. This involves dividing 67 years of quarterly data, from 1955 Q1 to 2022 Q2, into 140 overlapping segments, each spanning 30 years (120 quarters). The breakdown of this data into 140 30-year samples is visualized in Figure B.1. Each segment is used to estimate the model and generate out-of-sample forecasts. The mean squared error (MSE) for these forecasts is calculated by averaging the squared forecast errors across all segments at each forecast horizon. The results, showing the average out-of-sample MSEs across various forecast horizons, are presented in Figure 3.

B.3. *Data*

Figure B.2 displays the quarterly data used in this study, obtained and modified as follows, with percentage changes and gaps calculated using the log-difference method:

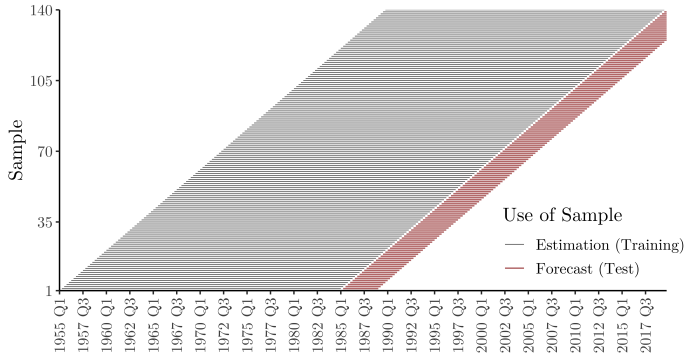


FIGURE B.1.—Sample

Note: This figure illustrates the division of data into estimation and forecasting periods for evaluating the out-of-sample performance of three local projection (LP) estimation methods, as illustrated in Figure 3 of Section 4. The quarterly data spans 67 years of U.S. macro indicators from 1955 Q1 to 2022 Q2, segmented into 140 overlapping samples. Each sample contains 30 years (120 quarters) of data. The y-axis represents the 140 distinct samples, chronologically ordered, and the x-axis shows the sample periods. The black lines indicate the estimation (training) periods, while the red lines denote the subsequent periods used for forecasting (testing).

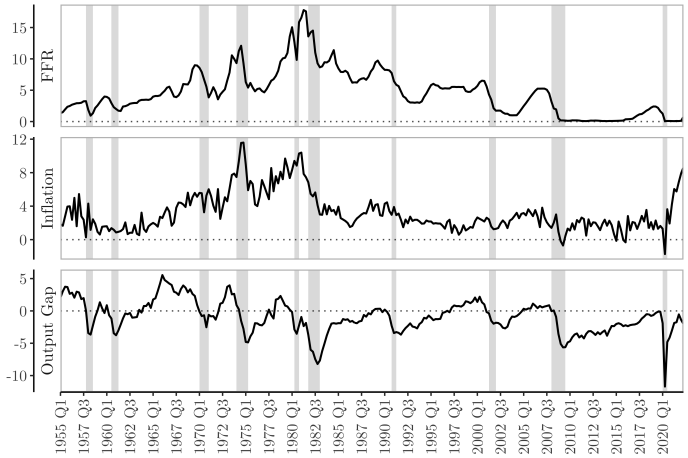


FIGURE B.2.—Data

Note: All variables are in annual percentage rates. The shaded areas represent recessions as dated by the National Bureau of Economic Research (NBER). Output gap is the annualized percentage gap between real and potential GDP of the United States, inflation is the annualized quarterly percentage change in the U.S. GDP chain-weighted price index, and FFR is the annualized quarterly average of the federal funds rate in percent. Percentage changes and gaps are calculated using the log-difference method.

1. *Output Gap*: This is the annualized percentage difference between real GDP and potential GDP. The real GDP data is provided by the U.S. Bureau of Economic Analysis (BEA) and accessed from the Federal Reserve Bank of St. Louis' FRED database us-

ing the ticker "GDPC1." Potential GDP data, sourced using the "GDPPOT" ticker, is provided by the U.S. Congressional Budget Office.

2. *Inflation*: This is calculated using the GDP chain-weighted price index, with data sourced from the BEA and obtained from FRED using ticker "GDPDEF." It is computed as the annualized quarterly percentage change in this price index.
3. *FFR*: The Federal Funds Rate, obtained form FRED using ticker "FEDFUNDS," is provided by the Board of Governors of the Federal Reserve System. This data reflects the annualized quarterly average of the federal funds rate, which is the interest rate at which depository institutions lend reserve balances to each other overnight.

APPENDIX C: PROOFS

This section provides proofs for Lemma 1 and Theorem 1. But first, the Frisch-Waugh-Lovell (FWL) Theorem B.1 is introduced, which is essential for these proofs.

C.1. *Frisch-Waugh-Lovell (FWL) Theorem for LPs and VARs*

Frisch and Waugh (1933) demonstrate that in a linear regression of a dependent variable on a trend and a set of explanatory variables, one can obtain the same coefficients by first regressing all variables on a trend, and then regressing the residuals of the dependent variable on the residuals of the explanatory variabes. Lovell (1963) generalizes this result by showing that this works for an arbitrary set of variables, not only for a trend. This generalized result is known as Frisch-Waugh-Lovell (FWL) Theorem. In this section, I state and prove the FWL Theorem for multivariate time series regressions so that it applies to local projections and vector autoregressions.

Consider the following multivariate time series regression:

$$\begin{aligned} Y_{t+h} &= P[Y_{t+h}|1, Y_t, \dots, Y_{t-p+1}] + u_{t+h} \\ &= AX_t + u_{t+h} \end{aligned} \tag{C.1}$$

with $X_t = \left[1\, Y_t' \cdots Y_{t-p+1}'\right]'$, where $\{Y_t\}_{t\in\mathbb{Z}}$ is a stochastic process of N-dimensional random vectors Y_t, P is the linear projection operator, A is an $N \times (1 + Np)$ matrix of coefficients, and u_{t+h} is an $N \times 1$ vector of orthogonal residuals, i.e. $E[u_{t+h}X_t'] = 0$, which is

well-defined under finite second moments. Note that for $h = 1, \ldots, H$, expression (C.1) is a local projection (1), and for $h = 1$, it is a vector autoregression (6).

Consider the sample version of (C.1) for a time series sample of T periods:

$$\begin{aligned} y_{t+h} &= \hat{P}[y_{t+h}|1, y_t, \ldots, y_{t-p+1}] + \hat{u}_{t+h} \\ &= \hat{A}x_t + \hat{u}_{t+h} \qquad\qquad t \in \mathcal{T} \subseteq \{p, \ldots, T-h\} \end{aligned} \tag{C.2}$$

with $x_t = \left[1\ y_t' \cdots y_{t-p+1}'\right]'$, where $\{y_t\}_{t\in\{1,\ldots,T\}}$ is a time series sample, $\hat{P}$ is the sample projection operator, $\hat{A}$ is the sample estimate of A, $\hat{u}_{t+h}$ is the sample residual, which sample correlation to x_t is zero, i.e. $\Sigma_{t\in\mathcal{T}}\hat{u}_{t+h}x_t' = 0$, and $\mathcal{T}$ is the set of time periods used for estimation.

The orthogonality conditions of (C.1) and (C.2) derive the following expression for population and sample parameters in terms of a stochastic process $\{Y_t\}_{t\in\mathbb{Z}}$ and a time series sample $\{y_t\}_{t\in\{1,\ldots,T\}}$:

$$\begin{aligned} E[u_{t+h}X_t'] = 0 \quad &\Rightarrow \quad AE[X_tX_t'] = E[Y_{t+h}X_t'] \\ \sum_{t\in\mathcal{T}} \hat{u}_{t+h}x_t' = 0 \quad &\Rightarrow \quad \hat{A}\left(\sum_{t\in\mathcal{T}} x_tx_t'\right) = \sum_{t\in\mathcal{T}} y_{t+h}x_t' \end{aligned} \tag{C.3}$$

and therefore; the parameters are unique if the matrices: $E[X_tX_t']$ and $\Sigma_{t\in\mathcal{T}}x_tx_t'$ are non-singular. Expression (C.3) estimates the LP in (1) for $h = 1, \ldots, H$, as well as the VAR in (6) for $h = 1$, where $\mathcal{T}$ is the set of time periods used for estimation. Note that minimizing the sum of squared residuals would derive the same parameters, which method is known as ordinary least squares (OLS).

LEMMA B.1—FWL Theorem for LPs and VARs: *Let W_t be a subset of X_t in (C.1), e.g. $X_t = \left[V_t'\ W_t'\right]'$. Then under finite second moments, i.e. $E\|Y_t\|^2 < \infty$, it is possible to partial out W_t as follows, in both, population and sample:*

$$\begin{aligned} Y_{t+h} - P[Y_{t+h}|W_t] &= A(X_t - P[X_t|W_t]) + u_{t+h} \\ y_{t+h} - \hat{P}[y_{t+h}|w_t] &= \hat{A}(x_t - \hat{P}[x_t|w_t]) + \hat{u}_{t+h} \end{aligned}$$

where $\hat{P}[y_{t+h}|w_t]$, and $\hat{P}[x_{t+h}|w_t]$ are estimated with the same sample $\mathcal{T}$ as $\hat{A}$ in (C.3).

PROOF: First, partial out W_t with the following multivariate regressions:

$$\begin{aligned} Y_{t+h} &= P[Y_{t+h}|W_t] + u^Y_{t+h} = FW_t + u^Y_{t+h} \\ X_t &= P[X_t|W_t] + u^X_{t+h} = GW_t + u^X_{t+h} \end{aligned} \tag{C.4}$$

where $E\left[u^Y_{t+h}W_t'\right] = 0$, and $E\left[u^X_{t+h}W_t'\right] = 0$. Second, substitute (C.1) into (C.4):

$$u^Y_{t+h} = (AG - F)W_t + Au^X_{t+h} + u_{t+h} \tag{C.5}$$

where W_t is orthogonal to u^Y_{t+h}, u^X_{t+h}, and u_{t+h}; and therefore, $AG - F = 0$, which proves Lemma B.1 in population. Third, estimate (C.4) with the same set of periods $\mathcal{T}$ as used to estimate $\hat{A}$ in (C.3), and substitute (C.2) into the sample version of (C.4):

$$\hat{u}^Y_{t+h} = (\hat{A}\hat{G} - \hat{F})w_t + \hat{A}\hat{u}^X_{t+h} + \hat{u}_{t+h} \qquad t \in \mathcal{T} = \{p, \dots, T-h\} \tag{C.6}$$

where w_t doesn't correlate with any of the remaining variables, i.e. $\Sigma_{t\in\mathcal{T}}\hat{u}_{t+h}w_t' = 0$, $\Sigma_{t\in\mathcal{T}}\hat{u}^Y_{t+h}w_t' = 0$, and $\Sigma_{t\in\mathcal{T}}\hat{u}^X_{t+h}w_t' = 0$; hence, $\hat{A}\hat{G} - \hat{F} = 0$, which proves Lemma B.1 for sample estimates, where all three regressions are estimated with the same sample $\mathcal{T}$. *Q.E.D.*

C.2. *Proof of Lemma 1*

This section provides the proof for Lemma 1, demonstrating that IRFs from single-equation local projections (SELPs) can be equivalently derived as IRFs from vector local projections (LPs). The proof employs the Frisch-Waugh-Lovell (FWL) Theorem B.1, which is applicable to both population and sample estimates. However, the FWL Theorem B.1 only applies to sample estimates when the parameters are estimated over the same set of time periods. To ensure consistency in the estimation process, the set of time periods $\mathcal{T}$ used for estimation is explicitly mentioned throughout the proof.

PROOF: Begin by expressing the SELP(p) model in (4) using the variable vector $Y_t = [V_t'\ x_t\ y_t\ W_t']'$. Here, $Y_{1:(K-1),t}$ denotes the first $(K-1)$ elements in Y_t, representing V_t,

$Y_{K,t}$ represents x_t, and $Y_{K+1,t+h}$ corresponds to y_t:

$$\begin{aligned} Y_{K+1,t+h} &= P[Y_{K+1,t+h}|1, Y_{1:(K-1),t}, Y_{K,t}, Y_{t-1}, \ldots, Y_{t-p}] + \nu^{(h)}_{t+h}(p) \\ &= \alpha^{(h)}(p) + \beta^{(h)}(p) Y_{K,t} + \delta^{(h)}(p)' Y_{1:(K-1),t} + \sum_{l=1}^{p} \lambda^{(h)}_{l}(p)' Y_{t-l} + \nu^{(h)}_{t+h}(p) \\ & \qquad h = 0, \ldots, H, \qquad \mathcal{T} = \{p+1, \ldots, T-H\} \end{aligned} \tag{C.7}$$

where P is the linear projection operator, $\mathcal{T}$ is the set of time periods used for estimation (C.3), T is the number of available time periods, and H is the maximum forecast horizon.

Applying the FWL Theorem B.1, the effects of contemporaneous variables in the SELP are isolated in (C.7):

$$\begin{aligned} Y_{K+1,t+h} &= P[Y_{K+1,t+h}|1, Y_{t-1}, \ldots, Y_{t-p}] \\ &\quad + \beta^{(h)}(p)(Y_{K,t} - P[Y_{K,t}|1, Y_{t-1}, \ldots, Y_{t-p}]) \\ &\quad + \delta^{(h)}(p)'(Y_{1:(K-1),t} - P[Y_{1:(K-1),t}|1, Y_{t-1}, \ldots, Y_{t-p}]) + \nu^{(h)}_{t+h}(p) \end{aligned} \tag{C.8}$$

Next, consider a vector LP(p) model as in (1) but lagged by one period:

$$\begin{aligned} Y_{t+h} &= b^{(h+1)}(p) + \sum_{l=1}^{p} B^{(h+1)}_{l}(p) Y_{t-l} + r^{(h+1)}_{t+h}(p) \\ & \qquad h = 0, \ldots, H, \qquad \mathcal{T} = \{p+1, \ldots, T-H\} \end{aligned} \tag{C.9}$$

The three linear projections in (C.8) use the same lag order and sample periods $\mathcal{T}$ as the $(K+1)$th equation of the linear projection in (C.9), allowing for substitution:

$$\begin{aligned} Y_{K+1,t+h} &= e'_{K+1} \left(b^{(h+1)}(p) + \sum_{l=1}^{p} B^{(h+1)}_{l}(p) Y_{t-l} \right) \\ &\quad + \beta^{(h)}(p) r^{(1)}_{K,t}(p) + \delta^{(h)}(p)' r^{(1)}_{1:(K-1),t}(p) + \nu^{(h)}_{t+h}(p) \end{aligned} \tag{C.10}$$

where $r^{(1)}_{t}(p)$ and LP(p) parameters are estimated in (C.9), and e'_{K+1} is an N-dimensional vector with zeros except for a one in the $(K+1)$th position, extracting the Kth row. Note

that the one-step ahead LP(p) residuals correspond to the VAR(p) residuals in (6), i.e., $r_t^{(1)}(p) = \xi_t(p)$, provided the VAR(p) is estimated using the same sample $\mathcal{T}$ defined in (C.9).

Perform a Cholesky decomposition of the VAR(p) residuals in (C.10):

$$r_t^{(1)}(p) = \xi_t(p) = \mathbf{A}\epsilon_t(p), \quad \mathbf{A}\mathbf{A}' = \text{Cov}(\xi_t(p)), \quad e_i'\mathbf{A}e_j = 0, \ \ \forall j > i \tag{C.11}$$

where $\mathbf{A}$ is a lower-triangular matrix that transforms $\xi_t(p)$ into uncorrelated shocks $\epsilon_t(p)$.

Given that $\mathbf{A}$ is lower-triangular, the residuals $r_t^{(1)}(p)$ in (C.10) can be replaced with the first K shocks $\epsilon_{1:K,t}(p)$ from (C.11):

$$\begin{aligned} Y_{K+1,t+h} &= e_{K+1}' \left(b^{(h+1)}(p) + \sum_{l=1}^{p} B_l^{(h+1)}(p) Y_{t-l} \right) \\ &+ \beta^{(h)}(p) e_K' \mathbf{A} e_K \epsilon_{K,t}(p) + \sum_{i=1}^{K-1} \left[\delta^{(h)}(p)' \ \beta^{(h)}(p) \ 0 \right] \mathbf{A} e_i \epsilon_{i,t}(p) + \nu_{t+h}^{(h)}(p) \end{aligned} \tag{C.12}$$

Next, consider an LP($p+1$) model as defined in (1) with an additional lag:

$$\begin{aligned} Y_{t+h} &= b^{(h)}(p+1) + \sum_{l=1}^{p+1} B_l^{(h)}(p+1) Y_{t+1-l} + r_{t+h}^{(h)}(p+1) \\ & h = 1, \ldots, H, \qquad \mathcal{T} = \{p+1, \ldots, T-H\} \end{aligned} \tag{C.13}$$

Using the FWL Theorem B.1, the effects of contemporaneous variables in the LP are isolated, and the resulting LP(p) residuals are substituted with their recursive decomposition in (C.11):

$$\begin{aligned} Y_{t+h} &= P[Y_{t+h}|1, Y_{t-1}, \ldots, Y_{t-p}] + B_1^{(h)}(p+1)(Y_t - P[Y_t|1, Y_{t-1}, \ldots, Y_{t-p}]) + r_{t+h}^{(h)}(p+1) \\ &= \left(b^{(h+1)}(p) + \sum_{l=1}^{p} B_l^{(h+1)}(p) Y_{t-l} \right) + B_1^{(h)}(p+1)\mathbf{A}\epsilon_t(p) + r_{t+h}^{(h)}(p+1) \end{aligned} \tag{C.14}$$

Comparing (C.12) with (C.14) reveals:

$$\beta^{(h)}(p) = \frac{e'_{K+1} B_1^{(h)}(p+1)\mathbf{A}e_K}{e'_K \mathbf{A} e_K} \qquad h = 0, \dots, H \tag{C.15}$$

where the numerator is the response of $Y_{K+1,t+h}$ to the shock $\epsilon_{K,t}(p)$, and the denominator is the response of $Y_{K,t}$ to the same shock. Thus, $\beta^{(h)}(p)$ represents an IRF to a shock normalized to have a unit impact on $Y_{K,t}$.

Comparing (C.15) with the definition of an LP IRF in (3) shows that the SELP(p) IRF in (5) is equivalent to the LP($p+1$) response to the VAR(p)-based impulse in (A.14), which proves Lemma 1. *Q.E.D.*

C.3. *Proof of Theorem 1*

This section proves Theorem 1, which states that LP and VAR models (1) and (6) are equivalent to their VAR and LP counterparts (14) and (19). The proof uses the FWL Theorem B.1, which applies to both, population and sample regressions; hence, the theorem applies to both, too. The FWL Theorem B.1 however only applies to sample estimates when the parameters are estimated with the same set of periods. Hence, in order to keep track of the set of periods used for estimation, I declare the set of time periods $\mathcal{T}$ used for estimation next to every equation. I first prove statement (a) that LPs are combinations of VARs and then statement (b) that VARs are combinations of LPs.

(a) LPs are Combinations of VARs

PROOF: Recall, a local projection (1) has the following form:

$$\begin{aligned} Y_{t+h} &= P[Y_{t+h}|1, Y_t, \dots, Y_{t-p+1}] + r_{t+h}^{(h)}(p) \qquad h = 1, \dots, H \\ &= b^{(h)}(p) + \sum_{s=1}^{p} B_s^{(h)}(p) Y_{t-s+1} + r_{t+h}^{(h)}(p) \qquad \mathcal{T} = \{p, \dots, T-H\} \end{aligned} \tag{C.16}$$

where P is the linear projection operator, $\mathcal{T}$ is the set of time periods used for estimation (C.3), T is the number of available time periods, and H is the maximum forecast horizon. Note that one could improve efficiency by maximizing the sample periods for every horizon h separately; that way, the sample ends at $T-h$ instead of $T-H$. I prove Theorem 1 for

samples that end at $T-H$, without loss of generality, because one could always change the maximum horizon to $H=1$, then $H=2$, etc.

In order to express a local projection as a combination of VARs, the LP(p) model (C.16) needs to fit inside a VAR model (6); hence, the lag length of the VAR has to increase with forecast horizon, as in (7):

$$Y_{t+h} = \gamma(p+h-1) + \sum_{l=1}^{p+h-1} \Gamma_l(p+h-1) Y_{t+h-l} + \xi_{t+h}(p+h-1) \qquad h=1,\dots,p+1, \quad \mathcal{T}=\{p,\dots,T-H\} \tag{C.17}$$

which are VAR($p+h-1$) models shifted forward by h periods, while the sample periods $\mathcal{T}$ remain fixed; hence, each VAR is estimated with a slightly different sample.

The FWL Theorem B.1 is used to partial out the LP(p) regressors in the VARs (C.17):

$$Y_{t+h} = P[Y_{t+h}|1, Y_t, \dots, Y_{t-p+1}] + \sum_{k=1}^{h-1} \Gamma_k(p+h-1)\Big(Y_{t+h-k} - P[Y_{t+h-k}|1, Y_t, \dots, Y_{t-p+1}]\Big) + \xi_{t+h}(p+h-1) \tag{C.18}$$

The linear projections in (C.18) use the same lag order and sample periods $\mathcal{T}$ as the local projections in (C.16), and thus can be substituted:

$$Y_{t+h} = b^{(h)}(p) + \sum_{s=1}^{p} B_s^{(h)}(p) Y_{t-s+1} + \sum_{k=1}^{h-1} \Gamma_k(p+h-1)\left(Y_{t+h-k} - b^{(h-k)}(p) - \sum_{s=1}^{p} B_s^{(h-k)}(p) Y_{t-s+1}\right) + \xi_{t+h}(p+h-1) \tag{C.19}$$

Rearrange (C.19) so that it mirrors a VAR:

$$Y_{t+h} = \left(b^{(h)}(p) - \sum_{k=1}^{h-1} \Gamma_k(p+h-1) b^{(h-k)}(p) \right) + \sum_{l=1}^{h-1} \Gamma_l(p+h-1) Y_{t+h-l} + \sum_{l=h}^{p+h-1} \left(B_{l-h+1}^{(h)}(p) - \sum_{k=1}^{h-1} \Gamma_k(p+h-1) B_{l-h+1}^{(h-k)}(p) \right) Y_{t+h-l} + \xi_{t+h}(p+h-1) \tag{C.20}$$

Since (C.20) has the same form, residuals, and sample periods $\mathcal{T}$ as the VAR$(p+h-1)$ models in (C.17), the following is true:

$$\begin{aligned} \Gamma_l(p+h-1) &= B_{l-h+1}^{(h)}(p) - \sum_{k=1}^{h-1} \Gamma_k(p+h-1) B_{l-h+1}^{(h-k)}(p) \qquad l = h, \ldots, p+h-1 \\ \gamma(p+h-1) &= b^{(h)}(p) - \sum_{k=1}^{h-1} \Gamma_k(p+h-1) b^{(h-k)}(p) \\ \xi_{t+h}(p+h-1) &= r_{t+h}^{(h)}(p) - \sum_{k=1}^{h-1} \Gamma_k(p+h-1) r_{t+h-k}^{(h-k)}(p) \qquad h = 1, \ldots, H \end{aligned} \tag{C.21}$$

which can be rearranged to match expression (15):

$$\begin{aligned} B_l^{(h)}(p) &= \sum_{k=1}^{p+h-1} \Gamma_k(p+h-1) B_l^{(h-k)}(p) \qquad l = 1, \ldots, p \\ b^{(h)}(p) &= \gamma(p+h-1) + \sum_{k=1}^{h-1} \Gamma_k(p+h-1) b^{(h-k)}(p) \\ r_{t+h}^{(h)}(p) &= \xi_{t+h}(p+h-1) + \sum_{k=1}^{h-1} \Gamma_k(p+h-1) r_{t+h-k}^{(h-k)}(p) \qquad h = 1, \ldots, H \end{aligned} \tag{C.22}$$

with normalization $B_l^{(1-k)}(p) = I_N$ for all $l = k \geq 1$, and $B_l^{(1-k)}(p) = 0$ otherwise.

Comparison of (C.22) with (15) reveals that $B_l^{(h)}(p) = \Psi_l^{(h)}(p)$, $b^{(h)}(p) = \psi^{(h)}(p)$, and $r_{t+h}^{(h)}(p) = \chi_{t+h}^{(h)}(p)$, which proves statement (a) of Theorem 1. Since the FWL Theorem B.1 applies to both, population regression and sample estimates, expression (C.22) also holds when the population parameters and residuals are replaced with sample estimates. *Q.E.D.*

(b) VARs are Combinations of LPs

PROOF: Recall, a vector autoregression (6) has the following form:

$$\begin{aligned} Y_t &= P[Y_t|1, Y_{t-1}, \dots, Y_{t-p}] + \xi_t(p) \\ &= \gamma(p) + \sum_{k=1}^{p} \Gamma_k(p) Y_{t-k} + \xi_t(p) \qquad \mathcal{T} = \{p+1, \dots, T\} \end{aligned} \tag{C.23}$$

where P is the linear projection operator, $\mathcal{T}$ is the set of time periods used for estimation (C.3), and T is the number of available time periods. Note that the only difference between an estimated VAR(p) and an estimated LP(p) at horizon $h = 1$ is that the latter does not include the last H periods.

In order to express a VAR as a combination of local projections, we need to fit LPs (1) inside the VAR (C.23). This can be done by partialling out some of the regressors in the VAR (C.23) by FWL Theorem B.1:

$$\begin{aligned} Y_t &= P[Y_t|1, Y_{t-h}, \dots, Y_{t-p}] \\ &\quad + \sum_{k=1}^{h-1} \Gamma_k(p) \Big(Y_{t-k} - P[Y_{t-k}|1, Y_{t-h}, \dots, Y_{t-p}] \Big) + \xi_t(p) \\ &\qquad\qquad h = 1, \dots, p+1, \quad \mathcal{T} = \{p+1, \dots, T\} \end{aligned} \tag{C.24}$$

where $P[Y_{t-k}|1, Y_{t-h}, \dots, Y_{t-p}]$ are LP($p-h+1$) predictions at horizon $h-k$. Since a VAR(p) consists of $p+1$ periods only: $\{Y_t, \dots, Y_{t-p}\}$, the local projection at horizon h cannot have more than $p-h+1 \leq p$ lags (C.23); hence, a VAR is a combination of equal- and lower-order LPs. Note that the set of time periods $\mathcal{T}$ used for estimation remains constant, while the time periods of the variables vary across projections; hence, the LPs are estimated with different sets of time periods.

It is necessary to add a second superscript when replacing the projection operators in (C.24) with LPs, because LPs of the same horizon $h-k$ are estimated with different sets

of time periods across different k:

$$
\begin{aligned}
Y_t = b^{(h,0)}(p-h+1) + \sum_{l=h}^{p} B_{l-h+1}^{(h,0)}(p-h+1)Y_{t-l} \\
+ \sum_{k=1}^{h-1} \Gamma_k(p) \left(Y_{t-k} - b^{(h-k,k)}(p-h+1) - \sum_{l=h}^{p} B_{l-h+1}^{(h-k,k)}(p-h+1)Y_{t-l} \right) + \xi_t(p)
\end{aligned}
\tag{C.25}
$$

where the second superscript is redundant in population, i.e. $b^{(i,j)}(s) = b^{(i,k)}(s)$, $B_l^{(i,j)}(s) = B_l^{(i,k)}(s)$, and $r_t^{(i,j)}(s) = r_t^{(i,k)}(s)$, for all k, but the sample estimates differ slightly because of distinct sample periods.

Rearrange (C.25) so that it mirrors a VAR:

$$
\begin{aligned}
Y_t = \left(b^{(h,0)}(p-h+1) - \sum_{k=1}^{h-1} \Gamma_k(p) b^{(h-k,k)}(p-h+1) \right) + \sum_{l=1}^{h-1} \Gamma_l(p) Y_{t-l} \\
+ \sum_{l=h}^{p} \left(B_{l-h+1}^{(h,0)}(p-h+1) - \sum_{k=1}^{h-1} \Gamma_k(p) B_{l-h+1}^{(h-k,k)}(p-h+1) \right) Y_{t-l} + \xi_t(p) \\
h = 1, \ldots, p+1, \quad \mathcal{T} = \{p+1, \ldots, T\}
\end{aligned}
\tag{C.26}
$$

Since (C.26) has the same form, residuals, and sample periods $\mathcal{T}$ as the VAR(p) model in (C.23), the following is true:

$$
\begin{aligned}
\Gamma_l(p) &= B_{l-h+1}^{(h,0)}(p-h+1) - \sum_{k=1}^{h-1} \Gamma_k(p) B_{l-h+1}^{(h-k,k)}(p-h+1) \qquad l = h, \ldots, p \\
\gamma(p) &= b^{(h,0)}(p-h+1) - \sum_{k=1}^{h-1} \Gamma_k(p) b^{(h-k,k)}(p-h+1) \\
\xi_t(p) &= r_t^{(h,0)}(p-h+1) - \sum_{k=1}^{h-1} \Gamma_k(p) r_{t-k}^{(h-k,k)}(p-h+1) \qquad h = 1, \ldots, p+1
\end{aligned}
\tag{C.27}
$$

where the second index in the superscript is redundant in population, and therefore, $\Gamma_l(p) = A_l(p)$, $\gamma(p) = a(p)$, and $\xi_t(p) = v_t(p)$, in (6) and (19), which proves statement (b) of Theorem 1. Since the FWL Theorem B.1 applies to both, population regression and sam-

ple estimates, expression (C.27) also holds when the population parameters and residuals are replaced with sample estimates. *Q.E.D.*

APPENDIX: REFERENCES

BLANCHARD, OLIVIER JEAN AND DANNY QUAH (1989): The Dynamic Effects of Aggregate Demand and Supply Disturbances, *The American Economic Review*, 79 (4), 655–673. []
[4, 7, 13, 34, 38, 39, Supp. 1, Supp. 7, Supp. 10, Supp. 11]

ENGLE, ROBERT F. AND CLIVE W. J. GRANGER (1987): Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, 55 (2), 251. [4, 7, 34, 38, Supp. 1]

FRISCH, RAGNAR AND FREDERICK V. WAUGH (1933): Partial Time Regressions as Compared with Individual Trends, *Econometrica*, 1 (4), 387–401. [Supp. 15]

FUHRER, JEFFREY C. AND GEORGE R. MOORE (1995a): Inflation Persistence, *The Quarterly Journal of Economics*, 110 (1), 127–159. [Supp. 12]

GALÍ, JORDI (1992): How Well Does the IS-LM Model Fit Postwar U.S. Data? *The Quarterly Journal of Economics*, 107 (2), 709–738. [Supp. 12]

HANSEN, PETER REINHARD (2005): Granger's Representation Theorem: A Closed-Form Expression for I(1) Processes, *The Econometrics Journal*, 8 (1), 23–38. [Supp. 9]

JOHANSEN, SØREN (1991): Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica*, 59 (6), 1551–1580. [7, 38, Supp. 1]

JORDÀ, ÒSCAR (2005): Estimation and Inference of Impulse Responses by Local Projections, *American Economic Review*, 95 (1), 161–182. [1, 2, 6, 8, 9, 20, 33, Supp. 12, Supp. 13]

LOVELL, MICHAEL C (1963): Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis, *Journal of the American Statistical Association*, 58 (304), 993–1010. [Supp. 15]

NUSBAUM, CHARLIE (2020): Aggregate Hours and Local Projections with Long-Run Restrictions, *SSRN Electronic Journal*, 1–31. [7, Supp. 10, Supp. 12]

RUDEBUSCH, GLENN AND LARS E. O. SVENSSON (1999): Policy Rules for Inflation Targeting, in *Monetary Policy Rules*, National Bureau of Economic Research, Inc, NBER Chapters, 203–262. [Supp. 12]