

© 2024 David Patrick Lundquist

FORECAST ADJUSTMENT UNDER SHOCKS: SIMILARITY-BASED SOLUTIONS TO  
UNPRECEDENTED EVENTS

BY

DAVID PATRICK LUNDQUIST

DISSERTATION

Submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in Statistics  
in the Graduate College of the  
University of Illinois Urbana-Champaign, 2024

Urbana, Illinois

Doctoral Committee:

Assistant Professor Gökçe Dayanıklı (Committee Member)  
Assistant Professor Daniel Eck (Chair and Advisor)  
Professor Jihyung Lee (Committee Member)  
Professor Feng Liang (Committee Member)

# Abstract

This work examines various forecasting challenges under conditionals that are not ideal. In the first chapter, we develop a procedure for forecasting the volatility of a time series immediately following a news shock by exploiting series that have experienced similar shocks. By positing news shocks to be affine functions of exogenous risk proxies, we are able to express particular news shocks as random realizations of a common data-generating process. In our model, random effects induce excess volatility. Those excess volatilities are then estimated as fixed effects and aggregated to adjust the GARCH volatility forecast of the time series under study by an additive term. The adjusted and unadjusted forecasts are evaluated using the unobservable but easily-estimated realized volatility (RV). Asymptotic results are established for our excess volatility estimators as well as our adjustment estimator. Numerical simulations are provided to illustrate conditions and hyperparameters under which our method thrives. A detailed real data application for forecasting volatility after the outcome of the 2016 United States presidential election demonstrates the applicability of the method.

The second chapter dramatically abstracts from the comfort of a particular model and instead probes situations in which one's default times series forecasting model must be adjusted, augmented, or abandoned completely, generating second-order questions about algorithmic design and the role of human judgment, as well as the information leveraged for forecast adjustment. This work further systematizes and unifies the rich landscape of model adjustment and model correction methods, with a special focus on forecast adjustment under the presence of news shocks, when unanticipated events may give an observer reason to doubt the credibility of the default forecasting function. We demonstrate the usefulness of similarity-based methods in forecasting and present a general framework dubbed Similarity-based Parameter Correction (SPC). We highlight several specific time series models that can benefit from SPC, along with formal results for some of those special cases.

We close with a thorough discussion of the state of forecasting under non-ideal conditions and the directions in which it may find the most success.

# Table of contents

# List of Abbreviations

DGP	Data-Generating Processes
GARCH	Generalized Conditional heteroskedasticity

# Chapter 1

# Post-Shock Volatility Forecasting Using Similarity-based Shock Aggregation

## 1.1 Introduction

Reacting to a seemingly unprecedented event might involve the question: what, if anything, does it resemble from the past? Such might be the case with event-driven investing strategies, where the identification of the event could arise via the news pages or corporate communications and hence contains a qualitative, narrative element [1]. Matching a current crisis to past events is a problem with unsurprising statistical angles: identification, sample size, weighting, risk, and robustness, among many others.

In the context of foreign exchange rate market structure, [2] speculate “[w]hether news is scheduled or non-scheduled its influence on exchange rates may be related to the state of the market at the time of the news arrival. News that arrives during periods of high uncertainty may have different effects on the exchange rate, than news that arrives in calmer periods.” The authors also note that non-scheduled news may require more time for markets to digest, leading to greater heterogeneity (including but not limited to higher dispersion) of responses. We take inspiration from these arguments, developing a method suitable to the conditions that typically accompany news shocks. Central to our endeavor is the observation that news shocks have both a qualitative (the description of the event as well as future contingents surrounding it) and quantitative component.

In this work, we focus on the second central moment of a time series, the volatility. One of the most important aspects of a positively-valued time series  $(P_t)_{t \in \mathbb{N}}$ , especially financial time series, is the volatility of the return series  $(r_t)_{t \in \mathbb{N}}$ . A financial asset’s price series may exhibit behavior that makes inapplicable and uninterpretable the traditional methods of time series analysis. In contrast, the return series is scale-free [3], easily-interpreted, and often at least weakly stationary. Our reasons for studying volatility are two-fold. Whereas once returns were thought to react to specific news events, now stock price movements are believed to be overwhelmingly noise-based (see [4] and references therein), and even if one could construct credible models for describing and forecasting price series and return series, that would not necessarily tell us much about the variability of such forecasts nor enlighten us about the evolution of the variability of  $(P_t)_{t \in \mathbb{N}}$  and

$(r_t)_{t \in \mathbb{N}}$  or under rapidly changing conditions.

No matter how a time series or its transformations are modeled, forecasting in the presence of news shocks requires a methodological framework that sensibly incorporates relevant information that has yet to manifest in market price or derivative quantities like volatility. In this setting, regime-change models (see [5] and citations therein) are of little use because under the assumption of a known exogenous shock, there is no need to estimate a regime-change time, nor is there data following the exogenous shock event to fit a model. Asymmetric GARCH models were an early attempt to account for the fact that negative returns typically beget greater volatility than positive returns [6]. Asymmetry and mixed-frequency are employed in [7] to forecast under the presence of extreme shocks. Problematically, any asymmetric model will depend upon the observation of a signed return to provide the most updated volatility forecast, but under the circumstances posited herein, no such return has been observed.

Similar problems and more exist for Realized GARCH models [6] in our setting. These models incorporate observable measures of volatility, known as “realized measures”, like implied volatility (IV). However, under the assumptions herein, no post-shock data is available, and even if it were, Realized GARCH does not promise to perform well, since Black-Scholes-implied volatility is a biased estimator of volatility [8], [9], with the bias increasing in times of crises, when options may be far out-of-the-money. GARCH models have been shown slow to adapt to spikes in volatility [10].

The approach herein sidesteps the modeling shortcomings of Realized GARCH and other approaches which require observations of post-shock outcomes by substituting modeling assumptions. The method herein proceeds under the assumption that similar news events occasion volatility shocks arising from a common conditional shock distribution. The procedure proposed does not require post-shock information like returns or market-implied quantities from the time series under study. Hence, we also avoid questions about what realized measure to use and when as well as questions about the usefulness of high-frequency data, although these remain intriguing avenues for future work.

The primary methodological tool presented in this work is fixed effect estimation followed by a distance-based weighting procedure for pooling those estimates. The use of fixed effect estimation for the study of structural shocks has a pedigree in macroeconomic analysis ([11] as referenced in [12]; see also discussion of deterministic exogenous events in [13]). We employ fixed effect estimation on the basis of a well-established conceptual assumption that shocks of economic time series can be modeled as mixtures, in particular, mixtures of ordinary innovations and shocks due to rare events (see [14] and references therein). In the forecasting literature, the term “intercept correction” has come to refer to a modeling technique in which nonzero errors are explicitly permitted [15], [16]. They summarize the literature as distinguishing two families of intercept correction: so-called “discretionary” intercept corrections that attempt to account for future events, without hard-coding an ad-hoc adjustment into the model specification, and second, “automated” intercept corrections that attempt to adjust for persistent misspecification using past errors. [17] use weighted subsets of a scalar series’ own past to correct forecast errors by an additive term. [18] introduce a similarity-based forecasting procedure for time-varying coefficients of a linear model. [19] employ a form of intercept correction in order to adjust forecasts to the COVID-19 shock in the spring of 2020 based on the proportional misses of the same model applied to the Great Recession.

A researcher interested in forecast adjustment can choose between procedures that discretionary or automated, a variety of choices for the collection of data assembled to perform the correction, whether the data is internal (i.e. from the time series itself) or external, the parametric term to be corrected (e.g. intercept, coefficients), if any, as well as the correction function (i.e. the mapping from the data to the

corrective term), including the weighting applied to the assembled data (e.g. Nearest-Neighbor, arithmetic mean, kernel methods).

Our procedure is a discretionary procedure for intercept correction that incorporates systematically data internal or external to the time series under study. The correction function, as we shall see, includes a convex combination of fixed effect estimates from a donor pool. In the causal inference literature, a donor pool is a set of units of observation that share distributional properties with the prime object of analysis. In [20], the authors build upon previous work in causal inference whereby a treatment effect can be estimated via comparison with a synthetic time series that represents the control. In their setting, the synthetic unit is constructed using a convex combination of the donors. The particular convex combination employed is a function of the distance between the time series under study and the donors. Donors that are closer to the time series under study are given greater weight. Conversely, donors that are unlike the time series under study receive little to no weight. [21] adapt the notion of a donor pool as well as distance-based weighting for the purpose of prediction. Their one-step-ahead forecasts use distance-based-weighting to aggregate shock estimates from the donor pool’s time series according to the similarity to the time series under study. Their approach does not take into account the ARCH effects commonly observed in time series, especially financial times series, leaving unaccounted for the variability in the variability of time series forecasts. Outside of [21], we know of no prior work that both introduces a parametric specification for nonzero errors and introduces a procedure for weighting appropriately the nonzero errors of similar shocks occurring outside the time series under study. Likewise, we are not familiar with any prior work that attempts to account for anticipated nonzero errors using an explicit parametric adjustment, i.e., what we will call a “correction function”.

The method proposed herein is supported by both numerical evidence as well as a real data example. The numerical evidence takes the form of simulations and analysis of the method’s performance across a considerable grid of varying parameters, allowing us to evaluate reasonable hypotheses about the method but also uncover some intriguing nuances. The real data example, on the other hand, demonstrates the usefulness of the method by applying it to the United States (US) presidential election. We find that the method would have allowed a researcher to predict the volatility of a financial asset that immediately followed the 2016 US presidential. Our discussion of this example sheds light on some subtleties of the method as well as its robustness to various choices that must be made to construct and assemble donors and appropriate covariates [22].

## 1.2 Setting

In order to motivate our procedure, we provide a visual illustration. In Figure ??, we show how the aggregation of estimated shock-induced excess volatilities from donors in the donor pool works when the correction function is a specially-chosen convex combination of fixed effects from the donor pool. Our method assumes a credible, parsimonious parameterization in which the shock is an affine transformation of several key covariates. The key intuition behind this shock parameterization is that as the strength of the linear signal increases relative to the idiosyncratic error, the GARCH estimation of these effects increases in accuracy. From this, it follows that the aggregated shock estimate increases in accuracy.

### 1.2.1 A Primer on GARCH

We define the log return of an asset between  $t - 1$  and  $t$  as  $r_t = \log(\frac{P_t}{P_{t-1}})$ , where  $P_t$  denotes the price at time  $t$ . The class of ARIMA( $p, d, q$ ) models [23] provides a framework for modeling the autoregressive structure of  $r_t$ .



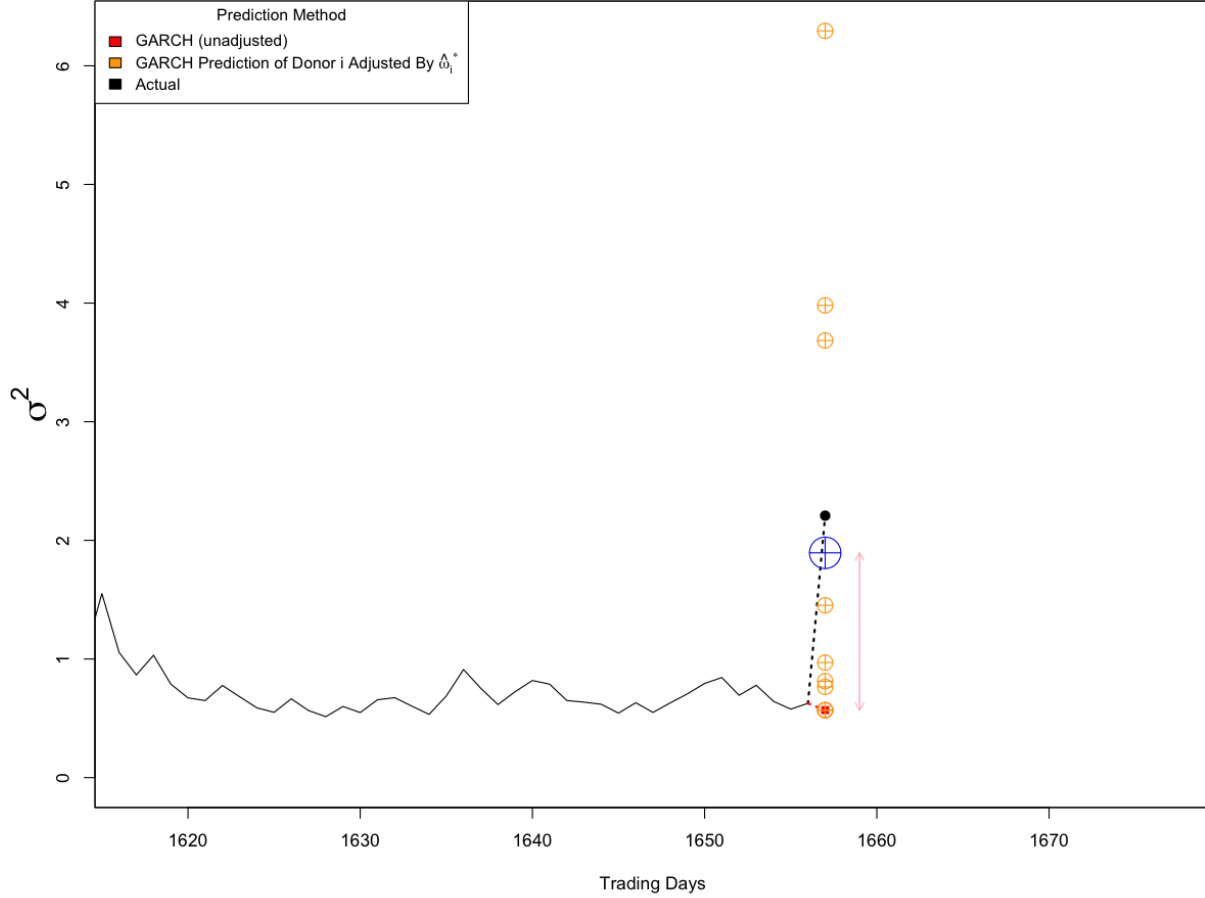


Figure 1.1: The simulated time series experiences a volatility shock between trading days 1,656 and 1,657. The GARCH prediction, in red, fails even to approach the volatility spike at  $T^* + 1$ , as do several adjusted predictions, which are orange. In contrast, the GARCH forecast adjusted by  $\hat{\omega}^* = \sum_{i=2}^{n+1} \pi_i \hat{\omega}_i^*$ , a convex combination of the estimated shocks in the donor pool, achieves directional correctness as well as a smaller absolute loss in its prediction. The pink vertical line serves to indicate the adjustment of size  $\hat{\omega}^*$  that allows the blue bullseye to approach more closely the ground truth.

These models assume a certain dependence structure between  $r_t$  and  $(r_k)_{k \leq t}$ , yet their errors — often called innovations in the financial time series context due to how they represent the impact of new information — are nevertheless assumed to be i.i.d. with mean zero and constant variance. The ARCH [24] and GARCH [25] models provide elegant alternatives to the homoskedasticity assumption. In fact, the GARCH framework in its most basic form disregards  $r_t$  and instead turns its interest to the series  $r_t^2$  (once properly centered, i.e. after assuming a mean-model for returns).

To that end, let  $a_t = r_t - \mu_t$ , where  $\mu_t$  is the (potentially time-varying) mean of the log return series  $r_t$ .

We thus derive a mean-zero process  $(a_t)_{t \in \mathbb{N}}$  with the property that  $\mathbb{E}[a_t^2] = \text{Var}[a_t]$ . Under the assumption of time-invariant volatility, the series  $a_t^2$  should exhibit no significant autocorrelation at any lag  $\ell \geq 1$ . This assumption motivates tests for ARCH effects, that is, tests for the clustering of volatility. These tests explore

the alternative hypothesis that  $\sigma_t^2$  is not only a time-varying parameter but furthermore a function of past squared residuals of the mean model. In particular, the ARCH( $m$ ) model is an autoregressive model in which  $\sigma_t^2$  is a deterministic function of the past  $m$  values of  $r_t^2$ . The GARCH( $m, s$ ) framework take this one step further by modeling  $\sigma_t^2$  as a linear combination of the past  $m$  values of  $r_t^2$  and well as the past  $s$  values of  $\sigma_t^2$ . In functional form, a GARCH process (sometimes called a strong GARCH process [[p. 19]francq2019garch]) is given by

$$\begin{aligned}\sigma_t^2 &= \omega + \sum_{k=1}^m \alpha_k a_{t-k}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \\ a_t &= \sigma_t \epsilon_t \\ \epsilon_t &\stackrel{iid}{\sim} E[\epsilon_t] = 0, Var[\epsilon_t] = 1 \\ \forall k, j, \alpha_k, \beta_j &\geq 0 \\ \forall t, \omega, \sigma_t &> 0 .\end{aligned}$$

Assuming further that  $\sigma_t^2$  depends on a vector of exogenous covariates  $\mathbf{v}_t$ , we have a GARCH-X( $m, s$ ). The volatility equation then becomes

$$\sigma_t^2 = \omega + \sum_{k=1}^m \alpha_k a_{t-k}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 + \gamma^T \mathbf{v}_t . \quad (1.1)$$

### 1.2.2 Model setup

We will suppose that a researcher has multivariate time series data  $\mathbf{y}_{i,t} = (r_{i,t}, \mathbf{v}_{i,t})$ ,  $t = 1, \dots, T_i$ ,  $i = 1, \dots, n+1$ , where  $r_{i,t}$  is scalar and  $\mathbf{v}_{i,t}$  is a vector of covariates such that  $\mathbf{v}_{i,t} | \mathcal{F}_{i,t-1}$  is deterministic. Suppose that the analyst is interested in forecasting the volatility of  $r_{1,t}$ , the first time series in the collection, which we will denote *the time series under study*. We require that each time series  $\mathbf{y}_{i,t}$  is subject to an observed news event following  $T_i^* \leq T_i + 1$  and before witnessing  $T_i^* + 1$ . We are implicitly leveraging the fact that financial assets are heavily-traded during market hours, yet only thinly traded (if traded at all) outside market hours. In contrast, the arrival of market-moving news does not obey any such restrictions. In light of the foregoing, we can denote our collection of GARCH-X volatility equations of interest using the following notation

$$\sigma_{i,t}^2 = \omega_i + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{v}_{i,t} .$$

Let  $I(\cdot)$  be an indicator function. Let  $T_i$  denote the time length of the time series  $i$  for  $i = 1, \dots, n+1$ , and let  $T_i^*$  denote the largest time index prior to the arrival of the news shock, with  $T_i^* < T_i$ , to ensure that there is at least one post-shock realization for each series  $i$ . Let  $\delta, \mathbf{v}_{i,t} \in \mathbb{R}^p$ . Let  $\mathcal{F}_i$  with a single-variable subscript denote a univariate, time-invariant  $\sigma$ -algebra, and  $\mathcal{F}_{i,t}$  denote the canonical product filtration for donor  $i$  at time  $t$ . Let  $D_{i,t}^{return} = I(t \in \{T_i^* + 1, \dots, T_i^* + L_{i,return}\})$  and  $D_{i,t}^{vol} = I(t \in \{T_i^* + 1, \dots, T_i^* + L_{i,vol}\})$ , and

let  $L_{i,return}, L_{i,vol}$  denote the lengths of the level and volatility shocks, respectively. For  $t = 1, \dots, T_i$  and  $i = 1, \dots, n + 1$ , the model  $\mathcal{M}_1$  is defined as

$$\begin{aligned}\sigma_{i,t}^2 &= \omega_i + \sum_{k=1}^{m_i} \alpha_{i,k} a_{i,t-k}^2 + \sum_{j=1}^{s_i} \beta_{i,j} \sigma_{i,t-j}^2 + \gamma_i^T \mathbf{v}_{i,t} + \omega_{i,t}^*, \\ \mathcal{M}_1: \quad a_{i,t} &= \sigma_{i,t}((1 - D_{i,t}^{return})\epsilon_{i,t} + D_{i,t}^{return}\epsilon_{i,t}^*), \\ \omega_{i,t}^* &= D_{i,t}^{vol}[\mu_{\omega^*} + \delta^T \mathbf{v}_{i,T_i^*+1} + u_{i,t}],\end{aligned}$$

with time-invariant error structure

$$\begin{aligned}\epsilon_{i,t} &\overset{iid}{\sim} \mathcal{F}_\epsilon \text{ with } \mathbb{E}_{\mathcal{F}_\epsilon}(\epsilon) = 0, \text{Var}_{\mathcal{F}_\epsilon}(\epsilon) = 1, \\ \epsilon_{i,t}^* &\overset{iid}{\sim} \mathcal{F}_{\epsilon^*} \text{ with } \mathbb{E}_{\mathcal{F}_{\epsilon^*}}(\epsilon) = \mu_{\epsilon^*}, \text{Var}_{\mathcal{F}_{\epsilon^*}}(\epsilon^*) = \sigma_{\epsilon^*}^2, \\ u_{i,t} &\overset{iid}{\sim} \mathcal{F}_u \text{ with } \text{Var}_{\mathcal{F}_u}(u) = \sigma_u^2, \\ \epsilon_{i,t} &\perp\!\!\!\perp \epsilon_{i,t}^* \perp\!\!\!\perp u_{i,t}.\end{aligned}$$

Let  $\mathcal{M}_0$  denote the subclass of  $\mathcal{M}_1$  models such that  $\delta \equiv 0$ . Note that  $\mathcal{M}_0$  assumes that nonzero shocks have no dependence on the covariates and are i.i.d. with  $\mathbb{E}[\omega_{i,t}^*] = \mu_{\omega^*}$ , where the lack of indices  $i$  or  $t$  on  $\mu_{\omega^*}$  indicates that it is shared across donors and is time-invariant. Models  $\mathcal{M}_1$  and  $\mathcal{M}_0$  parameterize changing dynamics after a shock is observed (when  $t \geq T_i^* + 1$  for each series  $i$ ). It is important to note that our modeling setup will additionally suppose that  $\mathbf{v}_{i,t}$  will be observed before  $\sigma_{i,t}^2$  is observed.

Note also the divergences from [21]. As already established, the model presented here is intended to account for ARCH effects. Additionally, here we also diverge by modeling the shocks  $\omega_{i,t}^*$  as time-varying quantities obeying a mixture distribution. In particular, prior to and after the shock, the  $\omega_{i,t}^*$  are uniformly zero. During the shock, i.e. from  $T_i^* + 1$  through  $T_i^* + L_i^{vol}$ , the  $\omega_{i,t}^*$  are distributed i.i.d, with a core signal component that depends solely on the unknown parameter  $\delta$  and the observable vector  $\mathbf{v}_{i,T_i^*+1}$ . Wherever necessary, to distinguish the two distributions that make up the shock distribution, we use the term *nonzero shocks* to refer to the shocks that are not uniformly zero.

The nonzero shocks each include an idiosyncratic noise term that varies across time and donors. This feature allows the nonzero shocks to exhibit some variation as the shock progresses, even within a single donor. The noise term could be modeled with additional complexity, including serial correlation, or could exhibit deterministic or stochastic decay throughout the shock period, proportional to a deterministic or stochastic function of time, or could peak somewhere in the middle of the shock period. Each of these ideas might correspond to some particular, realistic feature of the world.

The model  $\mathcal{M}_1$  is defined by a volatility equation and mean equation, as is any GARCH model. The decision to model the volatility shock  $\omega_{i,t}^*$  as an additive random effect is consistent with our setup. However, the decision to model the level effect  $\epsilon_{i,t}^*$  as a temporary rupture in the otherwise i.i.d. sequence of innovations  $\epsilon_{i,t}$  may not be intuitive. One way of arguing for this choice is that, in a discrete time series model, if we assume the arrival of news in the time between  $T_i^*$  and  $T_i^* + 1$ , we do not have an easy way to express a conditional distribution of the innovation  $\epsilon_{i,T_i^*+1}$  given the arrival of information. Using  $\epsilon_{i,t}^*$  thus breaks this impasse. This justification also explains why we do not parameterize the level shock at  $T_i^* + 1$  as a sum of two shocks,  $\epsilon_{i,T_i^*+1}$  and  $\epsilon_{i,T_i^*+1}^*$ , which would represent the level shock as generated by two independent sources of stochasticity. While we want to model the level shocks at  $T_i^* + 1$  as potentially large in absolute value, we also want to retain the property of a unitary source of noise.

## 1.3 Methodology for Similarity-based Parameter Correction

We now introduce how a researcher can carry out predictions via distance-based weighting. The first step is to gather the  $p$  covariates that will parametrize a change in dynamics to model  $\mathcal{M}_1$  and will represent the  $p$ -dimensional space in which weights are generated. This set of covariates parameterizes the shock to  $\sigma_{i,t}^2$  of magnitude  $\omega_{i,t}^*$  in model  $\mathcal{M}_1$  above. Let  $\mathbf{V}_t \in \mathbb{R}^{p \times n}$  be formed with columns  $\mathbf{v}_{i,t}$ ,  $i = 2, \dots, n+1$ . Without loss of generality, let the indices of each  $\mathbf{v}_{i,t}$  be shifted and arranged such that  $\mathbf{V}_{T^*+1}$  has as columns  $\mathbf{v}_{2,T_2^*+1}, \dots, \mathbf{v}_{n+1,T_{n+1}^*+1}$ . Formally, in vector notation, the donor shocks are given by  $\vec{\omega}_t = \vec{\mu}_{\omega^*} + \delta^T \mathbf{V}_t + \vec{u}_t$ , where we have suppressed the  $D_{i,t}^{vol}$  notation for simplicity. Ideally,  $\mathbf{V}_t$  will display ‘balance’ in that  $p$  covariates exist for each of the  $n$  donors. In practice, missing values, corrupted values, or unacceptably extreme or noisy estimates may necessitate some sort of matrix completion, a problem that we do not tackle in this work.

### 1.3.1 Forecasting

We now turn to the next section, where  $\mathbf{V}_t$  is employed in a procedure to arrive at a forecast adjustment. For illustration, we present two one-step-ahead forecasts for the time series under study. The first is the unadjusted GARCH forecast. The second is the adjusted forecast, which differs by the additive term  $\hat{\omega}^*$  that is computed via our distance-based weighting procedure. These forecasts are:

$$\begin{aligned} \hat{\sigma}_{unadjusted, T_1^*+1}^2 &= \hat{\mathbb{E}}[\sigma_{1, T_1^*+1}^2 | \mathcal{F}_{T_1^*}] &= \hat{\omega}_1 + \sum_{k=1}^{m_1} \hat{\alpha}_{1,k} a_{1, T_1^*+1-k}^2 + \sum_{j=1}^{s_1} \hat{\beta}_{1,j} \sigma_{1, T_1^*+1-j}^2 + \hat{\gamma}_1^T \mathbf{v}_{1, T_1^*+1}, \\ \hat{\sigma}_{adjusted, T_1^*+1}^2 &= \hat{\mathbb{E}}[\sigma_{1, T_1^*+1}^2 | \mathcal{F}_{T_1^*}] + \hat{\omega}^* &= \hat{\omega}_1 + \sum_{k=1}^{m_1} \hat{\alpha}_{1,k} a_{1, T_1^*+1-k}^2 + \sum_{j=1}^{s_1} \hat{\beta}_{1,j} \sigma_{1, T_1^*+1-j}^2 + \hat{\gamma}_1^T \mathbf{v}_{1, T_1^*+1} + \hat{\omega}^* . \end{aligned}$$

Note that GARCH models can be parameterized as ARMA models on the squares of the scalar time series  $a_{i,t}^2$  [3], [26], assuming that  $a_{i,t}$  satisfies fourth-order stationarity. This fact matters for forecasting because the  $h$ -step-ahead forecasting function for a GARCH model is, just like for an ARMA model, the conditional expectation function,  $\mathbb{E}[\sigma_{i, T_i^*+h}^2 | \mathcal{F}_{T_i^*}]$ , or practically speaking, the estimate thereof,  $\hat{\mathbb{E}}[\sigma_{i, T_i^*+h}^2 | \mathcal{F}_{T_i^*}]$  [27]. Here we have presented one-step-ahead forecasts for a GARCH-X( $m, s$ ). For  $h = 2, 3, 4, \dots$ , the conditional expectation is computed recursively, as is standard for iterative autoregressive forecasts. Note that  $h$ -step-ahead predictions will require, of course, some reasonable idea about the length of time for which the adjustment quantity  $\hat{\omega}^*$  should be included. Additionally, if  $\mathbf{v}_{1,t}$  is used as an exogenous regressor in order to estimate  $\gamma_1$  prior to the shock period, then strictly speaking,  $\mathbf{v}_{1, t+h} | \mathcal{F}_t$  must be deterministic — at least for the relevant  $(t, h)$  pair for which a forecast is sought. In practice, it may be acceptable for  $\mathbf{v}_{1, t+h} | \mathcal{F}_t$  to be merely well-estimated. Thus for simplicity of exposition, we will focus on the one-step-ahead forecast unless stated otherwise.

### 1.3.2 Excess Volatility Estimators

The problem of aggregating estimated donor shocks begins with the data constraints. Let us first introduce useful notation. Let  $\hat{\omega}_{i,*}^*$  denote the shock estimate for donor  $i$  that is obtained via fixed effect estimation over time points  $T_i^* + 1, \dots, T_i^* + L_{i,vol}$ . That estimation procedure is justified by the assumption that at each nonzero shock point, the shocks will differ but will be equal in distribution.

Taking the estimated shocks as a given, we observe the pair  $(\{\hat{\omega}_{i,*}^*\}_{i=2}^{n+1}, \{\mathbf{v}_i\}_{i=2}^{n+1})$ . Let  $\Delta^{n-1} = \{\pi \in \mathbb{R}^n : \sum_{i=1}^n \pi_i = 1, \pi_i \geq 0, i = 1, \dots, n\}$ . We wish to recover weights  $\{\pi_i\}_{i=2}^{n+1} \in \Delta^{n-1}$  leading to favorable forecasting properties. These weights are used to define and compute an aggregate shock estimate

$$\hat{\omega}^* = \sum_{i=2}^{n+1} \pi_i \hat{\omega}_{i,*}^*, \quad (1.2)$$

which will be taken to be our forecast adjustment term. Since the weights  $\{\pi_i\}_{i=2}^{n+1}$  are computed using  $\mathcal{F}_{T_i^*}$ , the set  $\{\pi_i\}_{i=2}^{n+1}$  is deterministic and observed, *modulo* any stochastic ingredient in the numerical methods employed to approximate  $\mathbf{v}_{1,T_1^*+1}$  using a convex combination of donor covariates. We say more about the properties of the shocks  $\omega_{i,T_1^*+1}^*, \dots, \omega_{i,T_i^*+L_i^{vol}}^*$  in section ??.

Following [28],[20],[21], let  $\|\cdot\|_{\mathbf{S}}$  denote any semi-norm on  $\mathbb{R}^p$ , and define

$$\{\pi\}_{i=2}^{n+1} = \arg \min_{\pi} \|\mathbf{v}_{1,T_1^*+1} - \mathbf{V}_{T^*+1}\pi\|_{\mathbf{S}}.$$

In the forecast combination literature, it is of interest whether the weights employed to aggregate forecasts strive toward and meet various optimality criteria [29], [30]. In our endeavor, there are at least two senses of optimal weights that one might be interested in. First, we can think of optimal weights as a set  $\{\pi_i\}_{i=2}^{n+1}$  such that  $\omega_1 = \sum_{i=2}^{n+1} \pi_i \hat{\omega}_{i,*}$ , i.e.,  $\omega_1$  is recovered perfectly, as it belongs to convex hull of the estimated shocks. However,  $\omega_1$  is never revealed to the practitioner, and hence there is no way of verifying the extent to which this condition is satisfied.

A more promising aim is finding weights such that  $\mathbf{v}_1 = \sum_{i=2}^{n+1} \pi_i \mathbf{v}_{i,T_1^*+1}$ , meaning that the covariates of the time series under study lies within the convex hull of the donor covariates. This condition underwrites asymptotic results in [20], and the intuition there extends to this work: if the shock is parameterized by an affine function of covariates, then finding a linear combination that recreates the shock should serve us well. Because the method proposed uses a point in  $\Delta^{n-1}$ , it is important to head-off possible confusion. What we are proposing is not a forecast combination method. What we are aggregating and weighting (not combining) are subcomponents of forecasts, not forecasts themselves. Moreover, from a broader perspective, forecast combination is an inapt term for what is being proposed here. First, the donor time series do not provide forecasts, nor would forecasts be needed for random variables that have already been realized. Second and more fundamentally, the theoretical underpinnings of forecast combination, while diverse [30], are distinct from the setting presumed in this work.

Supposing that we can define optimal weights, then uniqueness may still be a concern. [21] discuss sufficient conditions for uniqueness as well as the implications of non-uniqueness. [31] invoke the Carathéodory Theorem to argue for the sparseness of the weight vector. We make additional comments as well.  $(\mathbb{R}^n, \|\cdot\|)$  is a Chebyshev space, and hence for any element  $x$  and any convex set  $C \subset \mathbb{R}^n$ , there exists a unique element  $y \in C$  that minimizes  $\|x - y\|$ . However, the pre-image of  $y$  with respect to a particular operator and constraint set might not be unique. Let  $p', n'$  denote the number of linearly independent rows of  $\mathbf{V}_t$  and linearly independent columns of  $\mathbf{V}_t$ , respectively. Let  $\text{col}(\cdot)$  denote the column space of a matrix, and let  $\text{Conv}(\cdot)$  denote the convex hull of a set of column vectors. The following table is useful for categorization of when a perfect fit and uniqueness prevail.

	$\mathbf{v}_1 \in \text{Conv}(\text{col}(\mathbf{V}_t))$	$\mathbf{v}_1 \notin \text{Conv}(\text{col}(\mathbf{V}_t))$
$p' \geq n'$	Perfect fit; minimizer unique	Fit not perfect; minimizer unique
$p' < n'$	Perfect fit; minimizer not necessarily unique, Carathéodory Theorem applies	Fit not perfect; minimizer not necessarily unique, Carathéodory Theorem applies

It should be noted, however, that even when the minimizer is unique, finding that unique minimizer may require numerical methods. In the event that one’s preferred numerical optimization routine fails to arrive at a single solution on the simplex in repeated runs of the algorithm, some sort of forecast combination might be advised. We discuss forecast combination later in this paper in the context of robustifying against misspecification.

On the questions of convex geometry and convex optimization, comparison with least-squares estimation is also illustrative. Consider least-squares estimation for the  $p$ -vector  $\delta$  in the shocks  $\omega_{i,t}^*$ :

$$\hat{\omega}_{OLS}^* = \arg \min_{\delta} \|\hat{\omega}^* - \delta^T \mathbf{V}_{T^*+1}\|_2^2,$$

where  $\mathbf{V}_{T^*+1}$  could include an adjoined row  $\mathbf{1}_n$ , in order to estimate the locator parameter of the shock. One problem is that this optimization problem is an optimization problem over  $p$ -vectors  $\delta$  — over linear combinations of the covariates, whereas what we seek is an  $n$ -vector — a linear combination of donors. Additionally, there is no guarantee that  $\hat{\omega}_{OLS}^* = \hat{\delta}_{OLS}^T \mathbf{v}_{1,T^*+1}$  would perform poorly, but given the small number of donors supposed in our setting, it is risky. Last, and perhaps most challenging for the least-squares estimate,  $\hat{\delta}_{OLS}^T$  faces the usual requirement of more donors than linearly independent covariates, and that may not often be easy to satisfy in our setting.

### 1.3.3 Ground Truth Estimators

The time-varying parameter  $\sigma_t^2$  is a quantity for which even identifying an observable effect in the real world is far more challenging. In this work, we use a common estimator of the variance called realized volatility (RV), one which has the virtue of being “model-free” in the sense that it requires no modeling assumptions [32]. The realized variance itself can be decomposed into the sum of a continuous component and a jump component, with the latter being less predictable and less persistent [33], cited in [34], two factors that further motivate the method employed herein.

Suppose we examine  $K$  units of time, where each unit is divided into  $m$  intervals of length  $\frac{1}{m}$ . We adapt the notation of [35]. Let  $p_t = \log P_t$ , and let  $\tilde{r}(t, \frac{1}{m}) = p_t - p_{t-\frac{1}{m}}$ . Suppressing the  $i$  index, we estimate the variance the log return series using Realized Volatility of the  $K$  consecutive trading days that conclude with day  $t$ , denoted  $RV_t^{K,m}$ , using

$$RV_t^{K,m} = \frac{1}{K} \sum_{v=1}^{Km} \tilde{r}^2(v/m, 1/m),$$

where the  $K$  trading days have been chopped into  $Km$  equally-sized blocks.

Assuming that the  $K$  units  $\tilde{r}(t, 1) = p_t - p_{t-1}$  are such that  $\tilde{r}(t, 1) \stackrel{iid}{\sim} N(\mu, \delta^2)$ , it is easily verified that  $\mathbb{E}[RV^{K,m}] = \frac{\mu^2}{m} + \delta^2$ , which is a biased but consistent estimator of the variance. We will proceed using  $m = 77$ , corresponding to the 6.5-hour trading day chopped into 5-minute blocks, with the first block omitted in order to ignore unusual trading behavior at the start of the day. Other sensible choices for  $m$  can replace

our specification if desired.

### 1.3.4 Loss Functions

We are interested in point forecasts for  $\sigma_{1,T_1^*+h}^2 | \mathcal{F}_{T_1^*}$ ,  $h = 1, 2, \dots$ , the  $h$ -step ahead conditional variance for the time series under study. Let  $L^h$  with the subscripted pair {prediction method, ground truth estimator}, denote the loss function for an  $h$ -step-ahead forecast using a given prediction function and ground truth estimator. For example, the  $h$ -step-ahead MSE when forecasting from time  $t$  using our method and using Realized Volatility as the ground truth is

$$\text{MSE}_{\text{adjusted prediction, RV}}^h = (\hat{\sigma}_{\text{adjusted prediction, } t+h}^2 - \hat{\sigma}_{RV, t+h}^2)^2 .$$

Also of interest in absolute percentage error for an  $h$ -step-ahead forecast, defined as

$$\text{APE}_{\text{method, ground truth}}^h = \frac{|\hat{\sigma}_{\text{method, } t+h}^2 - \hat{\sigma}_{\text{ground truth, } t+h}^2|}{\hat{\sigma}_{\text{ground truth, } t+h}^2} .$$

Finally, we introduce the QL (quasi-likelihood) Loss [36]:

$$\text{QL}_{\text{method, ground truth}}^h = \frac{\hat{\sigma}_{\text{ground truth, } t+h}^2}{\hat{\sigma}_{\text{method, } t+h}^2} - \log \frac{\hat{\sigma}_{\text{ground truth, } t+h}^2}{\hat{\sigma}_{\text{method, } t+h}^2} - 1 .$$

What distinguishes QL Loss is that it is multiplicative rather than additive. This has benefits, both practical and theoretical. As [36] explain, the technical properties of the QL Loss allow researchers to compare forecasts across heterogeneous time series, whereas additive loss functions like MSE unfairly penalize forecasts made under market turbulence. For this reason and others, we proceed to evaluate the method, both in simulations and real data examples, using the QL loss.

## 1.4 Properties of Volatility Shocks and Shock Estimators

In this section we will provide estimation properties for our shock estimator (??) and our adjusted forecast. These results will be with  $m_i = s_i = 1$ , for all  $i = 1, \dots, n+1$  in model  $\mathcal{M}_1$  unless otherwise stated, i.e. a GARCH(1,1) model with additional parameterizations for shocks. Note that the dual level-volatility shock in  $\mathcal{M}_1$  has a marginal effect on the conditional variance  $\sigma_{i,t}^2$  that reflects the geometric decay of innovations in autoregressive models. As usual, assume  $\alpha + \beta < 1$ . Furthermore, assume that both the volatility shock and the level shock are of length one only, and consider a circumstance with no exogenous covariate  $\mathbf{v}_{i,t}$ , except in the conditional shock distribution. Assume also that  $r \geq 2$ , which is necessary in order to isolate the effects of the level shock  $\epsilon_{i,t}^*$ . Then

$$\begin{aligned} \sigma_{i,T_i^*+r+1}^2 | \mathcal{F}_{T_i^*+r} &= \omega_i + \alpha_i a_{T_i^*+r}^2 + \beta_i \sigma_{i,T_i^*+r}^2 \\ &= \omega_i + \alpha_i (\sigma_{i,T_i^*+r} \epsilon_{T_i^*+r})^2 + \beta_i \sigma_{i,T_i^*+r}^2 \\ &= \omega_i + \sigma_{i,T_i^*+r}^2 (\alpha_i (\epsilon_{T_i^*+r})^2 + \beta_i) . \end{aligned} \tag{1.3}$$

In Equation (??), observe that  $\omega_{i,T_i^*+1}^*$  and  $\epsilon_{i,T_i^*+1}^*$  each appear at most once, through the term  $\sigma_{T_i^*+r}^2$ . This might lead one to suspect geometric decay of the shocks  $\omega_{i,T_i^*+1}^*$  and  $\epsilon_{i,T_i^*+1}^*$ . Such a suspicion is easier to justify by examining the conditional expectation of the variance,  $\mathbb{E}[\sigma_{i,T_i^*+r+1}^2 | \mathcal{F}_{T_i^*+r}]$ , which also happens to be the principal forecasting tool for a GARCH model [27]. Indeed, if we assume unit variance for all  $\epsilon_{i,t}$  except, of course,  $\epsilon_{i,t}^*$ , then we have

$$\begin{aligned}\mathbb{E}[\sigma_{i,T_i^*+r+1}^2 | \mathcal{F}_{T_i^*+r}] &= \mathbb{E}[\omega_i + \alpha a_{T_i^*+r}^2 + \beta \sigma_{i,T_i^*+r}^2 | \mathcal{F}_{T_i^*+r}] \\ &= \omega_i + \mathbb{E}[\alpha(\sigma_{i,T_i^*+r} \epsilon_{T_i^*+r}^*)^2 | \mathcal{F}_{T_i^*+r}] + \beta \sigma_{i,T_i^*+r}^2 \\ &= \omega_i + \alpha \sigma_{i,T_i^*+r}^2 + \beta \sigma_{i,T_i^*+r}^2 \quad (\text{Due to the unit variance assumption}) \\ &= \omega_i + \sigma_{i,T_i^*+r}^2 (\alpha + \beta) .\end{aligned}$$

By repeated substitution, in conditional expectation, the shock belongs to the equivalence class  $\mathcal{O}((\alpha + \beta)^r)$ . We generalize this observation in the following proposition.

**Proposition 1.** *Let  $a_{i,t}$  be a mean-zero time series obeying a GARCH(1,1) specification with unit-variance errors, all prior to the arrival of a volatility shock of length  $L_i^{vol} \geq 1$  and level shock of length  $L_i^{return} \geq 1$  at some time  $T_i^* + 1$ . Then for any  $i$ ,  $1 \leq i \leq n + 1$ , and  $r$  such that  $r \geq \max\{L_i^{return}, L_i^{return}\} + 1$ ,*

$$\mathbb{E}[\sigma_{i,T_i^*+r+1}^2 | \mathcal{F}_{T_i^*+r}] = \omega_i + (\alpha + \beta) \sigma_{i,T_i^*+r}^2 .$$

In other words, for a GARCH(1,1), once two time points removed from the longest shock length, the volatility shock and level shock can be subsumed into one. However, prior to being two time points removed, there is no such guarantee. For example, one can take  $r = 1$  and level shock of length at least 1 to see that

$$\begin{aligned}\mathbb{E}[\sigma_{i,T_i^*+2}^2 | \mathcal{F}_{T_i^*+1}] &= \mathbb{E}[\omega_i + \alpha a_{T_i^*+1}^2 + \beta \sigma_{i,T_i^*+1}^2 | \mathcal{F}_{T_i^*+1}] \\ &= \omega_i + \mathbb{E}[\alpha(\sigma_{i,T_i^*+1} \epsilon_{T_i^*+1}^*)^2 | \mathcal{F}_{T_i^*+1}] + \beta \sigma_{i,T_i^*+1}^2 \\ &= \omega_i + \alpha \sigma_{i,T_i^*+1}^2 (\mu_{\epsilon^*}^2 + \sigma_{\epsilon^*}^2) + \beta \sigma_{i,T_i^*+1}^2 \\ &= \omega_i + \sigma_{i,T_i^*+1}^2 (\alpha(\mu_{\epsilon^*}^2 + \sigma_{\epsilon^*}^2) + \beta),\end{aligned}$$

where  $(\alpha(\mu_{\epsilon^*}^2 + \sigma_{\epsilon^*}^2) + \beta)$  may be greater than 1, permitting explosive behavior, at least in the short term. After both shocks have been exhausted, their influence disappears quickly. This short-memory effect has implications for the method being developed herein. First, there may be different risks associated with overestimating and underestimating level shock and volatility shock lengths. Therefore, estimation of effects among the donors should err on the side of underestimating, not overestimating, the length of the max shock, since overestimation of the shock length brings with it the risk of underestimating  $\omega_{i,t}^*$ . Second, a practitioner of the method needs some idea of how long the the respective shocks in the time series under study might last. There are couple of obvious strategies: take all the donors, and over all the donor shock lengths, take the minimum. Alternatively, one could take the maximum.



### 1.4.1 Consistency of the shock estimators

The estimators  $\hat{\omega}_{i,t}^*$  in (??) are central to our forecast adjustment strategy. Here we show that the quantities  $\hat{\omega}_{i,t}^*$  possess an important asymptotic property that bolsters that overall reliability of the method proposed herein.

**Proposition 2.** *Assume*

1. For each  $i$ ,  $1 \leq i \leq n$ ,  $\{a_{i,t}\}_{t=0,\dots,T_i}$  obeys a GARCH- $X(m,s)$ , as laid out in Equation (??), with volatility shocks found in  $\mathcal{M}_1$ , where  $T_i$  is the length of the  $i$ th series.
2. For each  $i$ ,  $\{\omega_{i,t}^*\}_{t=0,\dots,T_i}$  is potentially non-zero at  $\{T_i^*+1, \dots, T_i^*+L_i^{vol}\}$ ,  $\omega_{i,T_i^*+1}^* \equiv \dots \equiv \omega_{i,T_i^*+L_i^{vol}}^*$ , and zero otherwise, where the arrival of  $T_i^*$  is governed by a time-invariant distribution on  $\{a_{i,t}\}_{t=0,\dots,T_i-1}$ , and both the arrival and conclusion of the shock is observable by the researcher.
3. The conditions in Assumption 0 of [37] hold.

Then for any  $i$ ,  $1 \leq i \leq n+1$ , and for any  $r$ ,  $1 \leq r \leq L_i^{vol}$ ,  $\hat{\omega}_{i,T_i^*+r}^* \xrightarrow{p} \omega_{i,T_i^*+r}^*$  as  $t \rightarrow \infty$ . Additionally,  $\hat{\omega}_{i,*}^* \xrightarrow{d} \omega_{i,T_i^*+r}^*$  as  $t \rightarrow \infty$ , and if for all  $i$ ,  $1 \leq i \leq n+1$ ,  $u_{i,t} \equiv 0$  on  $\{T_i^*+1, \dots, T_i^*+L_i^{vol}\}$ , then  $\hat{\omega}_{i,T_i^*+r}^* \xrightarrow{p} \omega_{i,T_i^*+r}^*$ .

**Lemma 1.** Under assumption ??, for each  $i$ ,  $i = 1, \dots, n+1$ ,  $\{\omega_{i,t}^*\}_{t=0,\dots,T_i}$  is a strictly stationary series.

### 1.4.2 Consistency of the Conditional Forecast Function

Having established the consistency of the estimators  $\hat{\omega}_{i,T_i^*+r}^*$ , we extend that result to prove asymptotic properties of the conditional forecast function itself.

**Proposition 3.** *Assume*

1. All conditions listed in Proposition ??.
2. There exist weights  $\{\pi_i\}_{i=2}^{n+1}$  such that  $\mathbf{v}_{1,T_1^*} = \sum_{i=2}^{n+1} \pi_i \mathbf{v}_{i,T_i^*}$ .

Then for any  $r$ ,  $1 \leq r \leq L_1^{vol}$ ,  $\hat{\sigma}_{adjusted,T_1^*+r}^2 \xrightarrow{d} \sigma_{1,T_1^*+r}^2$  as  $t \rightarrow \infty$  in the donor pool, and if for all  $i$ ,  $1 \leq i \leq n+1$ ,  $u_{i,t} \equiv 0$  on  $\{T_i^*+1, \dots, T_i^*+L_i^{vol}\}$ , then  $\hat{\sigma}_{adjusted,T_1^*+r}^2 \xrightarrow{p} \sigma_{1,T_1^*+r}^2$ .

### 1.4.3 Asymptotic Loss

We now evaluate the loss and risk of our method under two scenarios: first, under arbitrary distribution of  $\sigma_{t+1}^2$ , and then second, under the assumption that the data-generating process is correctly specified. We proceed with the same notation for adjusted and unadjusted forecasts introduced in Section ?? . For a 1-step-ahead forecast of  $\sigma_{t+1}^2$  where  $t = T^*$ , consider the difference

$$\begin{aligned}
& QL(\hat{\sigma}_{unadjusted,t+1}^2, \sigma_{t+1}^2) - QL(\hat{\sigma}_{adjusted,t+1}^2, \sigma_{t+1}^2) \\
&= \left( \frac{\sigma_{t+1}^2}{\hat{\sigma}_{unadjusted,t+1}^2} - \log \frac{\sigma_{t+1}^2}{\hat{\sigma}_{unadjusted,t+1}^2} - 1 \right) - \left( \frac{\sigma_{t+1}^2}{\hat{\sigma}_{adjusted,t+1}^2} - \log \frac{\sigma_{t+1}^2}{\hat{\sigma}_{adjusted,t+1}^2} - 1 \right) \\
&= \frac{\sigma_{t+1}^2}{\hat{\sigma}_{unadjusted,t+1}^2} - \frac{\sigma_{t+1}^2}{\hat{\sigma}_{adjusted,t+1}^2} + \log \frac{\hat{\sigma}_{unadjusted,t+1}^2}{\hat{\sigma}_{adjusted,t+1}^2} \\
&= \frac{\sigma_{t+1}^2 (\hat{\sigma}_{adjusted,t+1}^2 - \hat{\sigma}_{unadjusted,t+1}^2)}{\hat{\sigma}_{adjusted,t+1}^2 \hat{\sigma}_{unadjusted,t+1}^2} + \log \frac{\hat{\sigma}_{unadjusted,t+1}^2}{\hat{\sigma}_{adjusted,t+1}^2}. \tag{1.4}
\end{aligned}$$

For simplicity, we work with a GARCH(1,1) that experiences a volatility shock at a single time point for which we would like to provide a point forecast. Then (??) can be expressed as

$$\frac{\sigma_{t+1}^2 \hat{\omega}_{t+1}^*}{\hat{\sigma}_{adjusted,t+1}^2 \hat{\sigma}_{unadjusted,t+1}^2} + \log \frac{\hat{\omega} + \hat{\alpha} a_t^2 + \hat{\beta} \sigma_t^2}{\hat{\omega} + \hat{\alpha} a_t^2 + \hat{\beta} \sigma_t^2 + \hat{\omega}_{t+1}^*}.$$

It is easily verified that as  $\hat{\omega}_{t+1}^* \rightarrow 0^+$ , the difference in the losses goes to zero. On the other hand, as  $\hat{\omega}_{t+1}^*$  becomes large, the difference in the losses turns negative, with the lesson being that  $\hat{\omega}_{t+1}^*$  must be in appropriate proportion to  $\sigma_{t+1}^2$  in order for the adjusted forecast to outperform the unadjusted forecast. This explains why it is so important to avoid using a naive adjustment estimator,  $\bar{\omega}^*$ , the arithmetic mean of the estimated shocks. We conclude this section with a broader result.

**Proposition 4.** *Assume the conditions in Propositions ?? and ??. Then as  $t \rightarrow \infty$  in the donor pool,*

$$QL(\hat{\sigma}_{unadjusted,T^*+1}^2, \sigma_{T^*+1}^2) - QL(\hat{\sigma}_{adjusted,T^*+1}^2, \sigma_{T^*+1}^2) \xrightarrow{d} \frac{\omega_{1,T^*+1}^*}{\sigma_{1,T^*+1}^2 - \omega_{1,T^*+1}^*} + \log \frac{\sigma_{1,T^*+1}^2 - \omega_{1,T^*+1}^*}{\sigma_{1,T^*+1}^2} \geq 0,$$

and if for all  $i, 1 \leq i \leq n+1$ ,  $u_{i,t} \equiv 0$  on  $\{T_i^* + 1, \dots, T_i^* + L_i^{vol}\}$ , the convergence is in probability. Hence, a correctly specified  $\mathcal{M}_1$  model will outperform the unadjusted forecast asymptotically.

## 1.5 Numerical Examples

In this section, we demonstrate the effectiveness of the proposed method using Monte Carlo simulations. All simulations will use  $\mathcal{M}_1$  volatility models and  $\mathcal{M}_0$  models on the returns, i.e.  $D_{i,t}^{return} \equiv 0$  for each  $i$  and each  $t$ . We first explain the parameters to be varied, the parameters that remain fixed, and the behavior we expect to observe. For our purposes, a parameter is defined to be fixed whenever it does not vary across any of the simulations performed, and conversely, a parameter is varying whenever it varies across at least two simulations performed. The overarching story told by the simulations is that our method performs well as the magnitude of the signal in the shock grows relative to the variance of the idiosyncratic noise term of the shock.

### 1.5.1 Fixed Parameters

Each simulation,  $i = 1, \dots, n+1$  is a GARCH(1,1) process of length  $T_i$  chosen randomly from the set of integers  $\{756, \dots, 2520\}$ , corresponding to approximately 3-10 years of daily financial data. We fixed the intercept of the processes at the *garchx* package default of  $\omega = .2$  RePEc:pra:mprapa:100301. We also use the values  $\alpha = .1, \beta = .82$ , corresponding to a GARCH process with greater influence from past values of  $\sigma_{i,t}^2$  than past values of  $a_{i,t}^2$ .

### 1.5.2 Varying Parameters

We vary exactly five parameters:  $\mu_V, \sigma_V, \mu_\delta, \mu_{\omega^*}, \sigma_u$ , each of which is critical to evaluating the method's responsiveness to changing conditions in the shock distribution.  $\mu_V, \sigma_V$  govern the elements of the covariates.  $\mu_\delta$  interacts with  $\mathbf{V}_t$  via the dot-product operation, of course.  $\mu_{\omega^*}$  is a location parameter for the volatility shock, and  $\sigma_u$  is the standard deviation of the volatility shock.

We add an important note about the parameter  $\mu_\delta$ , which governs a vector  $\delta$  of length  $p$  with elements that increase monotonically in proportion to  $1, \dots, p$ , after which they are scaled by the factor  $\frac{2 \cdot \mu_\delta}{p(p+1)}$ , so that a randomly selected entry of the vector has mean  $\mu_\delta$ . The heterogeneity of the entries in  $\delta$  is critical to simulating the plausible heterogeneity that will exist in the random effects structure. Absent a heterogeneous vector  $\delta$ , the shock would not vary systematically with regard to each variable in the covariate vector. Such a setup would fail to approximate the realities of financial markets.

### 1.5.3 Evaluative Framework and Monte Carlo Results

Consistent with our evaluative framework declared in ??, we compare adjusted and unadjusted forecasts using QL Loss, calculating the fraction of the simulations in which the adjusted forecast yields a QL Loss no smaller than that of the unadjusted forecast. The simulations fall into three broad categories, which we will denote **Signal and Noise** (displayed in Figure ??), **Interaction between Signal and Volatility Profile Means** (displayed in Figure ??), and lastly, **Interaction between Shock Intercept and Shock Noise** (displayed in Figure ??).

The first category plots two parameters of the shock equation,  $\mu_\delta$  and  $\sigma_u$ , in grids while increasing the mean of the covariate vectors  $\mathbf{v}$  over three plots. Recall that  $\delta$  is a parameter shared between donors, whereas each vector  $\mathbf{v}_{i,t}$  in  $\mathbf{V}_{i,t}$  is i.i.d. multivariate normal. In Figure ??, we see very little variation across the grid. However, in Figures ?? and ??, where  $\mu_v$  is increased, we observe an expected phenomenon: for any fixed column of the plots, the performance of the adjusted forecast improves with increasing  $\mu_\delta$ . A somewhat subtler phenomenon is also observed: for a fixed row of the grids in Figures ?? and ??, for  $\mu_\delta \geq .5$ , increasing  $\sigma_u$  leads to a decline in performance. However, this is not observed for smaller levels of  $\mu_\delta$ , where the signal is too weak.

The second category of simulations examines  $\mu_\delta$  and  $\mu_v$ , while increasing the noise  $\sigma_u$ . The row-wise and column-wise increases in performance that we witnessed in the first category largely hold in the second category of simulations, with one similar disclaimer: the relationship is strained under large levels of noise, like  $\sigma_u = 1$ .

Lastly, the third set of simulations takes a look at the role of the intercept of the conditional shock distribution,  $\mu_{\omega^*}$ , in particular, its interaction with  $\sigma_u$ . We can think of  $\mu_{\omega^*}$  as being the sole component of  $\mathbf{v}$  that has no variance, i.e. because its corresponding entry in the vector  $\mathbf{v}$  is the scalar 1. For larger values of  $\mu_{\omega^*}$ , the outperformance of the adjusted forecast declines with increases in  $\sigma_u$ , but this effect is absent for smaller values of  $\mu_{\omega^*}$ .

## Signal and Noise

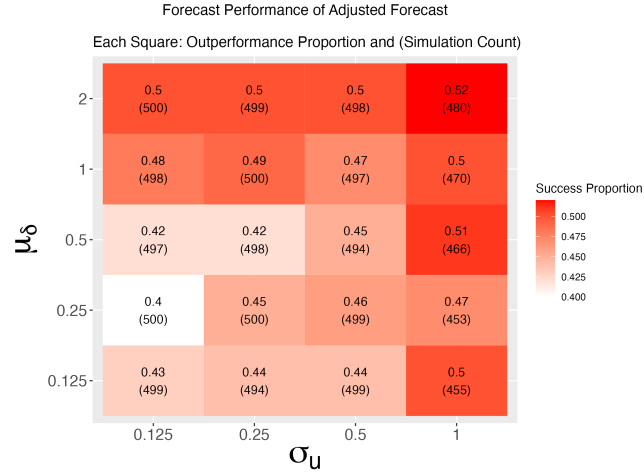


Figure 1.2: Fixed values:  $\mu_v = .125, \sigma_v = .125, \mu_{\omega^*} = .125$

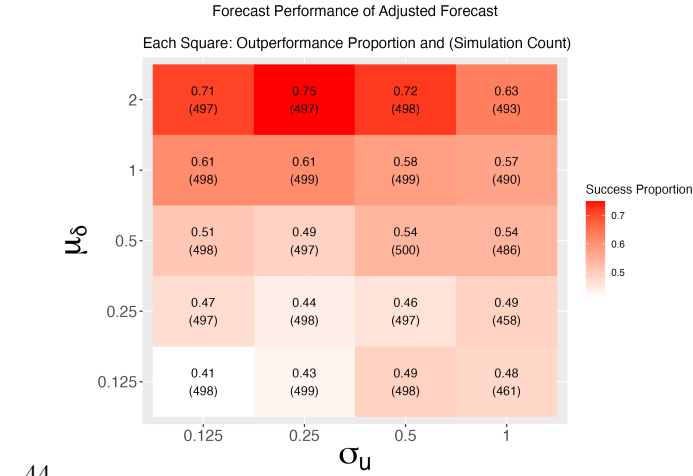


Figure 1.3: Fixed values:  $\mu_v = .5, \sigma_v = .125, \mu_{\omega^*} = .125$

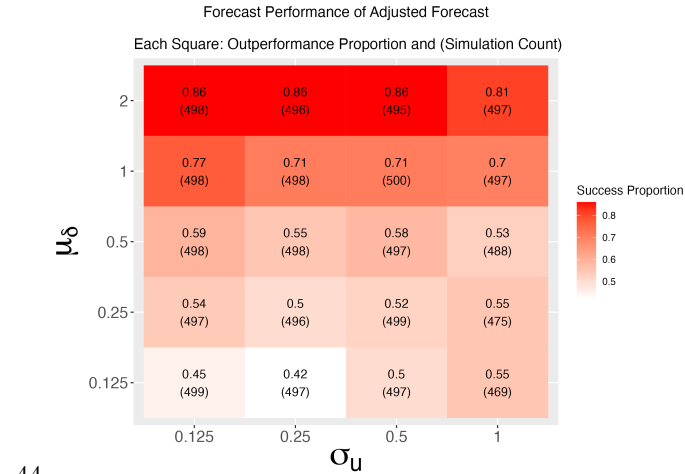


Figure 1.4: Fixed values:  $\mu_v = 1, \sigma_v = .125, \mu_{\omega^*} = .125$

Figure 1.5: In the progression from ?? to ?? to ??, we see that increasing  $\mu_\delta$  leads to improved performance of the adjusted forecast, and this effect is intensified as  $\mu_v$  increases. However, it is not consistently true that an increasing noise undermines the performance of the adjusted forecast. Rather, that phenomenon requires both large quantities of  $\mu_\delta$  and  $\mu_v$ .

## Interaction between Signal and Volatility Profile Means

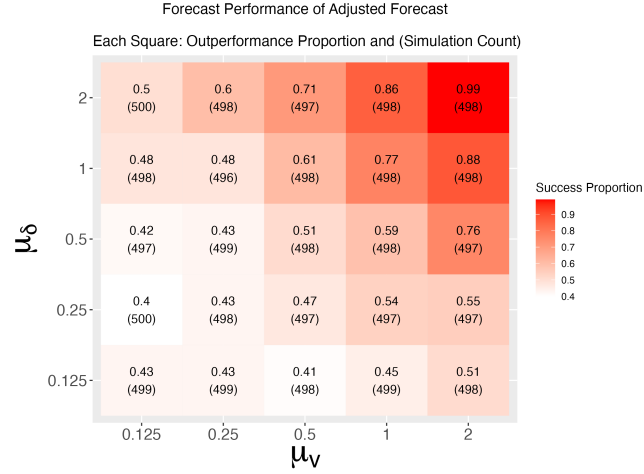


Figure 1.6: Fixed values:  $\sigma_v = .125$ ,  $\mu_{\omega^*} = .125$ ,  $\sigma_u = .125$

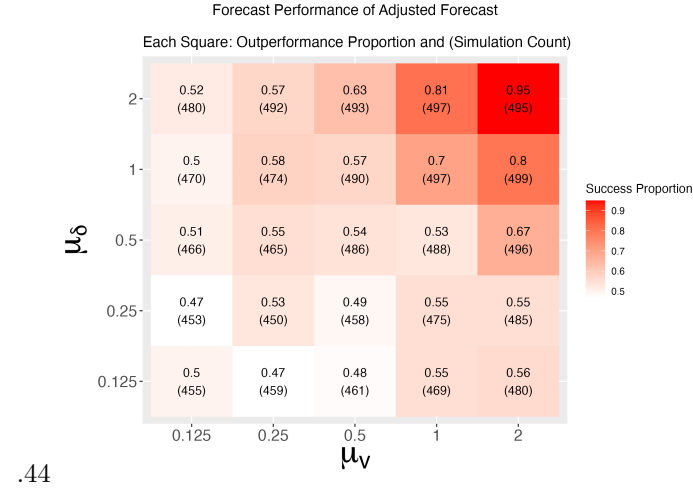
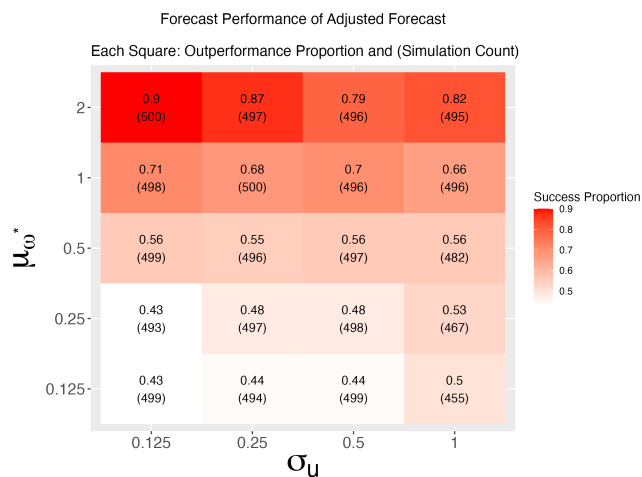


Figure 1.7: Fixed values:  $\sigma_v = .125$ ,  $\mu_{\omega^*} = .125$ ,  $\sigma_u = 1$

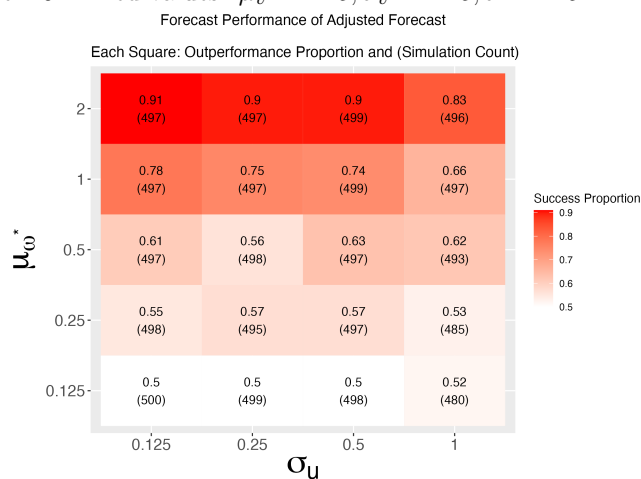
Figure 1.8: We compare the interaction of  $\mu_\delta$  and  $\mu_v$  at two different levels of shock noise. In the low-noise regime, the peak performance of the adjusted forecast is higher, and the ascent is faster along both dimensions. However, in the high-noise regime, the adjusted forecast performs well even at low levels of  $\mu_\delta$  and  $\mu_v$ .

## Interaction between Shock Intercept and Shock Noise



.44

Figure 1.9: Fixed values:  $\mu_v = .125, \sigma_v = .125, \delta = .125$



.44

Figure 1.10: Fixed values:  $\mu_v = .125, \sigma_v = .125, \delta = 2$

Figure 1.11: The interaction between the intercept  $\mu_{\omega}^*$  and  $\sigma_u$  suggests that the parameters behave as expected. However, a larger value of  $\delta$  in ?? attenuates the effect of increasing noise.

## Chapter 2

# Conclusion

We begin by recalling two salient quotes from important figures in the econometrics and forecasting literature:

[There exists] conflict between the intuitive notion that more relevant information should help in forecasting, and the hard reality that attempts to make it do so have not been uniformly successful [\[38\]](#)

incomplete information by itself is unlikely to play a key role in forecast failure (except if that information would forecast breaks). Consequently, using large amounts of data may not correct one of the main problems confronting forecasters, namely location shifts, unless that additional information is directly pertinent to forecasting breaks [\[39\]](#)

# References

- [1] W. Kenton, *Event-driven investing strategies and examples*, 2022. [Online]. Available: <https://www.investopedia.com/terms/e/eventdriven.asp>.
- [2] K. M. Dominguez and F. Panthaki, “What defines ‘news’ in foreign exchange markets?” *Journal of International Money and Finance*, vol. 25, no. 1, pp. 168–198, 2006.
- [3] R. S. Tsay, *Analysis of financial time series*. John wiley & sons, 2005, vol. 543.
- [4] J. Boudoukh, R. Feldman, S. Kogan, and M. Richardson, “Information, trading, and volatility: Evidence from firm-specific news,” *The Review of Financial Studies*, vol. 32, no. 3, pp. 992–1033, 2019.
- [5] L. Bauwens, A. Preminger, and J. V. Rombouts, “Regime switching garch models,” 2006.
- [6] P. R. Hansen, Z. Huang, and H. H. Shek, “Realized garch: A joint model for returns and realized measures of volatility,” *Journal of Applied Econometrics*, vol. 27, no. 6, pp. 877–906, 2012.
- [7] L. Wang, F. Ma, and G. Liu, “Forecasting stock volatility in the presence of extreme shocks: Short-term and long-term effects,” *Journal of Forecasting*, vol. 39, no. 5, pp. 797–810, 2020.
- [8] S. Mayhew, “Implied volatility,” *Financial Analysts Journal*, vol. 51, no. 4, pp. 8–20, 1995.
- [9] B. J. Christensen and N. R. Prabhala, “The relation between implied and realized volatility,” *Journal of financial economics*, vol. 50, no. 2, pp. 125–150, 1998.
- [10] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys, “Modeling and forecasting realized volatility,” *Econometrica*, vol. 71, no. 2, pp. 579–625, 2003.
- [11] C. D. Romer and D. H. Romer, “Does monetary policy matter? a new test in the spirit of friedman and schwartz,” *NBER macroeconomics annual*, vol. 4, pp. 121–170, 1989.
- [12] L. Kilian and H. Lütkepohl, *Structural vector autoregressive analysis*. Cambridge University Press, 2017.
- [13] R. F. Engle and A. J. Patton, “What good is a volatility model?” *Quantitative finance*, vol. 1, no. 2, p. 237, 2001.
- [14] R. F. Phillips, “Forecasting in the presence of large shocks,” *Journal of Economic Dynamics and Control*, vol. 20, no. 9-10, pp. 1581–1608, 1996.
- [15] D. Hendry and M. Clements, “On a theory of intercept corrections in macroeconometric forecasting,” 1994.
- [16] M. Clements and D. F. Hendry, *Forecasting economic time series*. Cambridge University Press, 1998.
- [17] P. Guerrón-Quintana and M. Zhong, “Macroeconomic forecasting in times of crises,” 2017.



- [18] Y. Dendramis, G. Kapetanios, and M. Marcellino, “A similarity-based approach for macroeconomic forecasting,” *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 183, no. 3, pp. 801–827, 2020.
- [19] C. Foroni, M. Marcellino, and D. Stevanovic, “Forecasting the covid-19 recession and recovery: Lessons from the financial crisis,” *International Journal of Forecasting*, vol. 38, no. 2, pp. 596–612, 2022.
- [20] A. Abadie, A. Diamond, and J. Hainmueller, “Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program,” *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 493–505, 2010.
- [21] J. Lin and D. J. Eck, “Minimizing post-shock forecasting error through aggregation of outside information,” *International Journal of Forecasting*, 2021.
- [22] S. Steegen, F. Tuerlinckx, A. Gelman, and W. Vanpaemel, “Increasing transparency through a multiverse analysis,” *Perspectives on Psychological Science*, vol. 11, no. 5, pp. 702–712, 2016.
- [23] G. Box, “Box and jenkins: Time series analysis, forecasting and control,” in *A Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century*, Springer, 2013, pp. 161–215.
- [24] R. F. Engle, “Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation,” *Econometrica: Journal of the econometric society*, pp. 987–1007, 1982.
- [25] T. Bollerslev, “Generalized autoregressive conditional heteroskedasticity,” *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [26] C. Francq and J.-M. Zakoian, *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2019.
- [27] E. Zivot, “Practical issues in the analysis of univariate garch models,” in *Handbook of financial time series*, Springer, 2009, pp. 113–155.
- [28] A. Abadie and J. Gardeazabal, “The economic costs of conflict: A case study of the basque country,” *American Economic Review*, vol. 93, no. 1, pp. 113–132, 2003.
- [29] A. Timmermann, “Forecast combinations,” *Handbook of economic forecasting*, vol. 1, pp. 135–196, 2006.
- [30] X. Wang, R. J. Hyndman, F. Li, and Y. Kang, “Forecast combinations: An over 50-year review,” *International Journal of Forecasting*, vol. 39, no. 4, pp. 1518–1547, 2023.
- [31] A. Abadie and J. Vives-i-Bastida, “Synthetic controls in action,” *arXiv preprint arXiv:2203.06279*, 2022.
- [32] T. G. Andersen and L. Benzoni, “Stochastic volatility,” *CREATES Research Paper*, no. 2010-10, 2010.
- [33] T. G. Andersen, T. Bollerslev, and F. X. Diebold, “Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility,” *The review of economics and statistics*, vol. 89, no. 4, pp. 701–720, 2007.
- [34] G. De Luca *et al.*, “Forecasting volatility using high-frequency data,” *Statistica Applicata*, vol. 18, 2006.
- [35] T. G. Andersen and L. Benzoni, “Realized volatility, working paper 2008-14,” 2008.
- [36] C. T. Brownlees, R. F. Engle, and B. T. Kelly, “A practical guide to volatility forecasting through calm and storm,” *Available at SSRN 1502915*, 2011.

- [37] H. Han and D. Kristensen, “Asymptotic theory for the qmle in garch-x models with stationary and nonstationary covariates,” *Journal of business & economic statistics*, vol. 32, no. 3, pp. 416–429, 2014.
- [38] M. P. Clements and D. F. Hendry, “Guest editors’ introduction: Information in economic forecasting,” *Oxford Bulletin of Economics & Statistics*, vol. 67, no. S1, pp. 713–753, 2005.
- [39] J. L. Castle, M. P. Clements, and D. F. Hendry, “Forecasting by factors, by variables, by both or neither?” *Journal of Econometrics*, vol. 177, no. 2, pp. 305–319, 2013.