C.1 Least Mean Square Filter

As discussed in the text, the variance paths of the structural shocks may be an object of interest in their own right. Some estimation methods produce these directly; MCMC draws values for the paths in order to conduct its numerical integration, and GARCH-based methods imply a filtered path. For others, like GMM, the estimated nuisance matrix instead provides limited information on the moments of the variance process. In these cases, a filtering algorithm is needed to obtain estimates of the variance paths, ideally one that requires few additional assumptions.

The Wiener filter is non-parametric, based only on moments of the measurement and the signal. Since the predictions are linear, this filter of course can produce negative variance estimates. This discussion follows that of Oppenheim & Verghese (2010) Chapter 11 and Brown & Hwang (1996), and primarily considers the Finite Impulse Response (FIR) version. The reader should consult the reference for a more detailed treatment. The Wiener filter has two basic assumptions:

**Assumption C.**

1. Signal $s_t$ and noise $v_t$ are stationary for all $t = 1, 2, \ldots, T$,
2. $s_t$ and $v_t$ have known autocorrelation and cross-correlation.

For this discussion, I strengthen these conditions in Assumption C’ for simplicity.

**Assumption C’.** For all $t = 1, 2, \ldots, T$ :

1. The signal is $\sigma_t^2$ with $\text{diag}(\sigma_t^2) = E[\varepsilon_t^2_t | \sigma_t]$,
2. The noise is $\text{vec}(\varepsilon_t^2 - \Sigma_t)$ (where $\Sigma_t = \text{diag}(\sigma_t^2)$),
3. $\sigma_t^2$ and $(\varepsilon_s^2 - \Sigma_s)$ are independent (a SV model),
4. $\varepsilon_t$ is independently normally distributed with $E[\Sigma_t^{-1/2}\varepsilon_t^2\Sigma_t^{-1/2}] = I_n$.

Assumption C’.3 implies that C.2 can be satisfied based on the data, without *a priori* knowledge of the signal or noise’s moments or training data on $s_t$. Assumption C’.4 will be exploited later. The Wiener filter minimizes the mean-squared error of the linear prediction

$$s_t = \sum_{j=0}^{J} a_j x_{t-j}$$
where \( x_t \) and \( s_t \) are vectors of the same length and \( a_j \) is a square matrix of the same dimension. \( J \) is the maximum lag considered in the prediction problem. Taking expectations and first order conditions results in the Wiener-Hopf equations,

\[
\begin{bmatrix}
R_x(0) & R_x(1) & \cdots & R_x(J) \\
R_x(1) & R_x(0) & \cdots & R_x(J-1) \\
\vdots & \vdots & \ddots & \vdots \\
R_x(J) & R_x(J-1) & \cdots & R_x(0)
\end{bmatrix}
\begin{bmatrix}
a_0 \\
a_1 \\
\vdots \\
a_J
\end{bmatrix}
= \begin{bmatrix}
R_{xs}(0) \\
R_{xs}(1) \\
\vdots \\
R_{xs}(J)
\end{bmatrix},
\]

where \( R_x(j) = E \left[ x(t) x(t+j)' \right] \) and \( R_{xs} = E \left[ x(t) s(t+j)' \right] \), or, more compactly,

\( R_x = aR_{xs} \).

The elements of \( R_x \) are easily estimated, as \( x_t \) is the input. However, it remains to calculate \( R_{xs} \). Now it is necessary to apply the assumptions of \( C \) and \( C' \). First, observe that in general, the input data is \( x_t = \text{vech} \left( \hat{\varepsilon}_t \hat{\varepsilon}_t' \right) \), where \( \hat{\varepsilon}_t = \hat{H}^{-1} \eta_t \). However, the off-diagonal elements of \( \hat{\varepsilon}_t \hat{\varepsilon}_t' \) are irrelevant for the linear prediction of both \( \hat{\varepsilon}_u \hat{\varepsilon}_u', u \neq t \) and \( \Sigma_u \forall u \). In other words,

\[
E[\varepsilon_{kt} \varepsilon_{lt} \varepsilon_{iu} \varepsilon_{mu}] = 0, k \neq l, t \neq u
\]

\[
E[\varepsilon_{kt} \varepsilon_{lt} \Sigma_u] = 0, k \neq l, t \neq u
\]

by \( C'.3 \) and \( C'.4 \). Therefore, specialize \( x_t = \text{diag} \left( \hat{\varepsilon}_t \hat{\varepsilon}_t' \right) \), and estimate \( R_x \) using the obvious sample counterpart. Now, turn to the estimation of \( R_{xs} \). By assumptions \( C'.3 \) and \( C'.4.4 \), \( R_{xs}(j) = R_x(j) \) for \( j = 1, 2, \ldots, J \). It remains to estimate \( R_{xs}(0) \). Again under assumptions \( C'.3 \) and \( C'.4 \),

\[
E \left[ \text{diag} \left( \hat{\varepsilon}_t \hat{\varepsilon}_t' \right) \text{diag} \left( \hat{\varepsilon}_t \hat{\varepsilon}_t' \right) \right] =
\begin{bmatrix}
3E \left[ \sigma_{1t}^2 \right] & E \left[ \sigma_{1t}^2 \sigma_{2t}^2 \right] & \cdots & E \left[ \sigma_{1t}^2 \sigma_{nt}^2 \right] \\
E \left[ \sigma_{2t}^2 \sigma_{1t}^2 \right] & 3E \left[ \sigma_{2t}^2 \right] & \cdots & E \left[ \sigma_{2t}^2 \sigma_{nt}^2 \right] \\
\vdots & \vdots & \ddots & \vdots \\
E \left[ \sigma_{nt}^2 \sigma_{1t}^2 \right] & E \left[ \sigma_{nt}^2 \sigma_{2t}^2 \right] & \cdots & 3E \left[ \sigma_{nt}^2 \right]
\end{bmatrix}.
\]

Thus, to obtain the desired matrix, \( E \left[ \sigma_t^2 \sigma_t^{2'} \right] \), the diagonal of an estimator of the above matrix can simply be divided by one-third. Thus, estimators of both \( R_x \) and \( R_{xs} \) are easily available under assumptions \( C \) and \( C' \). Finally, obtain \( \hat{a} = \hat{R}_x^{-1} \hat{R}_{xs} \). This requires that \( \hat{R}_x \) be invertible. Asymptotically, this will be the case since \( R_x \) is a covariance matrix. Then the filtered path for \( \sigma_t^2 \) is obtained as

\[
\hat{\sigma}_t^2 = \sum_{j=0}^{t-1} \hat{a}_j \text{diag} \left( \hat{\varepsilon}_t \hat{\varepsilon}_t' \right)_{t-j}.
\]

While the above discussion is highly specialized given assumption \( C' \), it is possible to relax \( C' \) and obtain similar results using more algebra.
C.2 Infill asymptotic implementations

Infill asymptotics examine the behavior of estimators as the frequency of observation increases, as opposed to the length of time spanned by the observations. In an infill setting (see e.g. Cressie (1993), Section 5.8 or Dahlhaus (2012), Section 2 for an introduction), an expectation can be consistently estimated over a finite time span as observations are taken over an increasing density of intervals. Infill asymptotic arguments are well-suited to non-stationary time series, (where standard asymptotics do not hold), that can be approximated by a stationary process in some neighborhood of each point in time. Foster & Nelson (1996), provide an application to covariance estimation. Given that the identification result of Theorem 2.1 makes no assumption of stationarity, the possibility of estimation approaches which also avoid such assumptions is appealing. The discussion below is intended as a non-technical overview; the interested reader should consult the references noted.

Kernel method

If a noisy time series, like the reduced-form variances considered here, is locally stationary, kernel smoothing can help eliminate the noise by smoothing values across a small window, see e.g. Hastie, Tibshirani, & Friedman (2009), Chapter 6. The objective is to estimate the regression function \( E_t [\eta_t \eta_t' | \eta_{1:T}] \) via

\[
\overline{\eta_t \eta_t'} = \sum_{N_{b_T}(t)} R \left( \eta_t \eta_t' \right),
\]

where \( R (\cdot) \) is a weighting function. More concretely, for a symmetric kernel \( k (l; b_T) \),

\[
\overline{\eta_t \eta_t'} = \frac{1}{b_T} \sum_{l=-b_T}^{b_T} k (l; b_T) \eta_{t-l} \eta_{t+l},
\]

where \( b_T \) is the bandwidth (I follow the discussion of local covariance estimation in Dahlhaus (2012), Section 2.2, with the exception that I subsume his \( b_T T \) into simply \( b_T \)). The consensus in the statistics literature is that the choice of kernel is relatively unimportant, driven by the high relative efficiency of many kernels relative to the “optimal” Epanechnikov kernel, see e.g. Silverman (1990), pg. 43. For the purposes of this paper, the Epanechnikov kernel is used, defined as

\[
k_{EP} (l; b_T) = \begin{cases} 
\frac{3}{4} \left( 1 - \left( \frac{l}{b_T} \right)^2 \right) & |l/b_T| \leq 1 \\
0 & \text{otherwise,}
\end{cases}
\]

see Hastie, Tibshirani, & Friedman (2009), pg. 193. The bandwidth used in the simulation study is \( b_T = 12 = \lambda T \), corresponding to a window twelve months either side of an observation. Experimentation with the bandwidth on the empirically calibrated AR(1) SV DGP shows very similar performance for values of \( b_T \) ranging from 6 to 24, with the minimum mean square error occurring at 12. When the data is non-stationary, the choice of bandwidth will be constrained to ensure the neighborhood of smoothing is locally stationary.
Applying a kernel smoothing algorithm estimates the path \( \widehat{\eta_t \eta_t'} \) from \( t = b_T + 1 \) to \( T - b_T \). Proposition 3 of Sentana & Fiorentini (2001) shows that under very general conditions, \( H \) is identified from such a path. This presents a very high-dimensional overidentified minimum distance problem – \( (n^2 + n)/2 \) parameters in \( H \) and \( (T - 2b_T)n \) structural variances.

The asymptotic properties of kernel estimates are discussed in detail in Theorem 3 in Dahlhaus (2012). To summarize, under regularity and strong smoothness conditions, Dahlhaus shows that estimates are asymptotically normal with a rate of \( (\lambda T)^{-1/2} \). The estimates also exhibit bias depending on the degree of non-stationarity of the true DGP. Under weaker conditions, slower rates are possible. Dahlhaus’ result is quite general and also covers the convergence of functions of the smoothed estimates, like the minimum distance problem considered here.

In a simple 2-dimensional model, it is possible to greatly reduce the dimensionality of the resulting minimum distance problem by avoiding the need to directly estimate the nuisance matrix, the two structural variances for each observation. The standard minimum distance set-up with three moments for the unique elements of \( \eta_t \eta_t' \) in each time period is asymptotically equivalent to minimum distance on the difference between the off-diagonal elements of \( H^{-1} \eta_t \eta_t' H^{-1} \) and zero. Given the nature of the problem, first estimating a smoothed path and then performing highly non-linear minimum distance on that path, there do not appear to be established methods for inference in this setting.

### Blocking method

A second approach retains the flavour of Rigobon’s identification argument. Essentially, if at least local stationarity is assumed within intervals of fixed length, the mean of \( \zeta_t \) can be estimated over blocks of data; the estimates will be consistent as \( \Delta t \), the time increment, tends to zero. Then, based on these subsamples, \( H \) is estimated using some minimum distance formulation following Proposition 3 of Sentana & Fiorentini (2001).

This has the downside that it is sensitive to the sub-sample length, as suggested by unreported simulations, but it is not susceptible to the main criticism leveled at Rigobon’s approach in Section 2.4, since arbitrarily-divided sub-samples of fixed length will not result in systematic non-diagonality. The expectation over each sub-sample remains fixed and generally differs across sub-samples even if the process is stationary, as the length of the sub-sample does not need to increase to apply the infill argument to estimators. Within each of these blocks, say \([t, \bar{t})\), estimate \( \int_t^{\bar{t}} E_t |H \Sigma_t H'| dt \) via \( \hat{\zeta}_{t,\bar{t}} \equiv \text{vech} \left( \frac{1}{T} \sum_{t=1}^{T} H \Sigma_t H' \right) \). From this “path” of discrete segments, \( H \) can be estimated. If the process is in fact stationary, as the length of the sub-sample does not need to increase to apply the infill argument to estimators. Within each of these blocks, say \([t, \bar{t})\), estimate \( \int_t^{\bar{t}} E_t |H \Sigma_t H'| dt \) via \( \hat{\zeta}_{t,\bar{t}} \equiv \text{vech} \left( \frac{1}{T} \sum_{t=1}^{T} H \Sigma_t H' \right) \). From this “path” of discrete segments, \( H \) can be estimated. If the process is in fact stationary, as detailed extensively in Jacod & Protter (2012), convergence to \( E \left[ H \Sigma_t H' | \xi \right] \) occurs for each block of length \( \lambda T \) at the standard rate \( (\lambda T/\Delta t)^{-1/2} \). For example, if \( \eta_t \) can be approximated on \([t, \bar{t})\) as an Itô semi-martingale (and some regularity conditions hold), their most basic CLT, Theorem 5.1.2, applies. If only local stationarity is assumed, the convergence results noted above for kernel estimators apply under the conditions noted in Dahlhaus (2012). As with the kernel method, it remains to estimate \( H \) from the overidentified system (assuming there are more than two blocks). Given these estimated innovation covariances for each block of data, a minimum distance estimator can be used to find the optimal \( H \) to satisfy the
overidentified system of equations.

C.3 Paths for the calibration of the DGPs

Figures C.1 and C.2 display sample paths for the variances for the empirically calibrated specifications and the “weak” specifications used to generate the simulation study, for the AR(1) SV and GARCH(1,1) processes respectively. Note that for each DGP, the “weak” paths show much smaller fluctuation about the mean, and that the difference in scale of fluctuation between empirical and weak is comparable across DGPs.

C.4 Additional simulation results

These tables and figures present additional simulation results discussed in the text. For Study 1, Table C.1 presents results for the AR(1) SV DGP. Figures C.3-C.5 present histograms for all estimators for both DGPs. For Study 2, histograms are included in Figures C.6-C.12 for estimates of both $H$ parameters for all DGPs and estimators, expanding on Table 2.5 of summary statistics in the main text.

<table>
<thead>
<tr>
<th>Table C.1: Median estimates for SV DGP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>norm</td>
</tr>
<tr>
<td>trace</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\eta_2^2$</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Median estimates of estimates for Rigobon-type estimators on the empirically-calibrated $AR(1)$ SV DGP, $T = 200$, 5,000 draws. The window indicates the length of the rolling window over which variances were compared to form subsamples. The norm indicates the method used to evaluate the magnitude of the variance over each window. The threshold indicates the value a window had to surpass for its central observation to be considered “high variance”. Estimation via the Sims (2014) method. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm. Since the RMSE must account for error in multiple parameter estimates, the MSE is computed for each, and then normalized by the square of the true parameter, before the root of the sum is taken.
Figure C.1: Comparison of sample variance paths for the log AR(1) SV process for empirical and “weak” calibrations. In the top panel, the paths are calibrated based on a bivariate SVAR(12) of the first factor of McCracken & Ng’s FRED-MD dataset (excluding FFR) and the FFR. In the lower panel, the paths are for the weak calibration, which divides the variance innovation covariance matrix by 10.
Figure C.2: Comparison of sample variance paths for the GARCH(1,1) process for empirical and “weak” calibrations. In the top panel, the paths are calibrated based on a bivariate SVAR(12) of the first factor of McCracken & Ng’s FRED-MD dataset (excluding FFR) and the FFR. In the lower panel, the paths are for the weak calibration, which divides the ARCH parameters by 1.5.
Figure C.3: Distribution of estimates given knowledge of the true Markov switching dates ("oracle"), $T = 200$, 5,000 draws. Estimation via the Sims (2014) method. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm.

Figure C.4: Distribution of estimates of estimates for Rigobon-type estimators on the empirically-calibrated Markov-switching DGP, $T = 200$, 5,000 draws. The window indicates the length of the rolling window over which variances were computed to form subsamples. The norm indicates the method used to evaluate the magnitude of the variance over each window. The threshold indicates the value a window had to surpass for its central observation to be considered "high variance". Estimation via the Sims (2014) method. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm.
Figure C.5: Distribution of estimates of estimates for Rigobon-type estimators on the empirically-calibrated AR(1) DGP, $T = 200$, 5,000 draws. The window indicates the length of the rolling window over which variances were computed to form subsamples. The norm indicates the method used to evaluate the magnitude of the variance over each window. The threshold indicates the value a window had to surpass for its central observation to be considered “high variance”. Estimation via the Sims (2014) method. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm.
Figure C.6: Distribution of estimates of $H_{21}$ and $H_{12}$ for various estimators for Markov switching DGP, $T = 200$, 5,000 draws. Details of estimators can be found in Table 4 of the main text. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm.
Figure C.7: Distribution of estimates of $H_{21}$ and $H_{12}$ for various estimators for empirically-calibrated GARCH(1,1) DGP, $T = 200$, 5,000 draws. Details of estimators can be found in Table 4 of the main text. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm.
Figure C.8: Distribution of estimates of $H_{21}$ and $H_{12}$ for various estimators for “weak” GARCH(1,1) DGP, $T = 200$, 5,000 draws. Details of estimators can be found in Table 4 of the main text. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm.
Figure C.9: Distribution of estimates of $H_{21}$ and $H_{12}$ for various estimators for empirically-calibrated AR(1) SV DGP, $T = 100$, 5,000 draws. Details of estimators can be found in Table 4 of the main text. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm.
Figure C.10: Distribution of estimates of $H_{21}$ and $H_{12}$ for various estimators for empirically-calibrated AR(1) SV DGP, $T = 200$, 5,000 draws. Details of estimators can be found in Table 4 of the main text. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm.
Figure C.11: Distribution of estimates of $H_{21}$ and $H_{12}$ for various estimators for empirically-calibrated AR(1) SV DGP, $T = 400$, 5,000 draws. Details of estimators can be found in Table 4 of the main text. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm.
Figure C.12: Distribution of estimates of $H_{21}$ and $H_{12}$ for various estimators for “weak” AR(1) SV DGP, $T = 200$, 5,000 draws. Details of estimators can be found in Table 4 of the main text. Labeling proceeds via an infeasible method matching $H$ estimates to the true $H$ to minimize $L_2$ norm.