9 Functional CLT. Fixed-b asymptotics.

9.1 Fixed bandwidth approach

The choice of truncation lag G in the Newey-West method is arbitrary. There are many ways of choosing this lag optimally, see Andrews (1991) for an example.
Kiefer, Vogelsang and Bunzel (2000) show that the accuracy of the tests based on the Newey-West variance estimator may be quite poor in finite samples, specifically tests over-reject the null (the estimated variance is 'too small'). They proposed an alternative where G is chosen such that $b \equiv \frac{G+1}{T} \to 0$ as $T \to \infty$. b is known as the bandwidth, and is kept fixed. For example, when G+1=T, b is fixed at 1. Under this approach $\hat{V}$ converges to a limiting random matrix that is proportional to V. The distribution of HAC robust tests based on $\hat{V}$ don't depend on the model's parameters (i.e. the distribution is pivotal), and can be tabulated.

Definition 9.1.1: Long-run variance

Sum of all the variances and covariances of a process, i.e. $\text{Var}(\sum_{t=1}^{T} \varepsilon_t)$.

Consider the simple regression on only a constant term

$$Y_t = \beta + \varepsilon_t.$$

The OLS estimator of β is $\hat{\beta}_{OLS} = \bar{Y}$, and under serial correlation:

$$Var(\hat{\beta}_{OLS}) = \frac{1}{T} V_T = \frac{1}{T}\left(\mathbb{E}\varepsilon_t^2 + \sum_{\ell=1}^{T-1} \frac{T-\ell}{T} 2\mathbb{E}(\varepsilon_t \varepsilon_{t-\ell})\right) \neq \frac{1}{T} Var(\varepsilon_t)$$

where the first equality follows from the previous lecture. As $T \to \infty$, the variance of the OLS estimator converges to the long-run variance of ε_t.
Newey-West

$$\hat{V}_{NW} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 + \sum_{\ell=1}^{G} \frac{G+1-\ell}{G+1} \frac{2}{T} \sum_{t=1+\ell}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-\ell}$$

KVB
KVB obtains an inconsistent estimator of V_T with $G = T-1$:

$$\hat{V}_{KVB} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 + \sum_{\ell=1}^{T-1} \frac{T-\ell}{T} \frac{2}{T} \sum_{t=1+\ell}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-\ell}$$

Here $\hat{\varepsilon}_t = Y_t - \bar{Y}$.

We can show that $\hat{V}_{KVB}$ is positive semi-definite as follows:

$$\begin{aligned}
\hat{V}_{KVB} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 + \sum_{\ell=1}^{T-1}\frac{T-\ell}{T}\frac{2}{T}\sum_{t=1+\ell}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-\ell} \\
&= \frac{1}{T}\mathbf{1}'\left(\frac{1}{T}\sum_{t=1+|i-j|}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_{t-|i-j|}\right)\mathbf{1} \quad \text{where 1 is a T-vector of ones} \\
&= \frac{1}{T^2}\mathbf{1}'ZZ'\mathbf{1} \\
\text{where } \underbrace{Z}_{T\times(2T-1)} &= \begin{bmatrix} \hat{\varepsilon}_1 & \hat{\varepsilon}_2 & \cdots & \hat{\varepsilon}_T & 0 & \cdots & 0 \\ 0 & \hat{\varepsilon}_1 & \cdots & \hat{\varepsilon}_{T-1} & \hat{\varepsilon}_T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\varepsilon}_1 & \hat{\varepsilon}_2 & \cdots & \hat{\varepsilon}_T \end{bmatrix}
\end{aligned}$$

Thus $\hat{V}_{KVB}$ is positive semi-definite. Further, since $\sum_{t=1}^{T}\hat{\varepsilon}_t = 0$ (property of OLS residuals) the sum of elements in the T-th column is zero. Moreover, the sum of elements in the $T+i$-th column gives:

$$\begin{aligned}
\sum_{t=i+1}^{T}\hat{\varepsilon}_t &= \sum_{t=0}^{T}\hat{\varepsilon}_t - \sum_{t=0}^{i}\hat{\varepsilon}_t \\
&= 0 - \sum_{t=0}^{i}\hat{\varepsilon}_t
\end{aligned}$$

Thus,

$$\mathbf{1}'Z = \left(\sum_{i=1}^{1}\hat{\varepsilon}_i, \sum_{i=1}^{2}\hat{\varepsilon}_i, \cdots, \sum_{i=1}^{T-1}, 0, -\sum_{i=1}^{1}\hat{\varepsilon}_i, -\sum_{i=1}^{2}\hat{\varepsilon}_i, \cdots, -\sum_{i=1}^{T-1}\hat{\varepsilon}_i\right)$$

Hence,

$$
\begin{aligned}
\hat{V}_{KVB} &= \frac{1}{T^2}\mathbf{1}'ZZ'\mathbf{1} \\
&= \frac{1}{T^2}\begin{bmatrix} \sum_{i=1}^{1}\hat{\varepsilon}_i & \cdots & \sum_{i=1}^{T-1}\hat{\varepsilon}_i & 0 & -\sum_{i=1}^{1}\hat{\varepsilon}_i & \cdots & -\sum_{i=1}^{T-1}\hat{\varepsilon}_i \end{bmatrix}\begin{bmatrix} \sum_{i=1}^{1}\hat{\varepsilon}_i \\ \vdots \\ \sum_{i=1}^{T-1}\hat{\varepsilon}_i \\ 0 \\ -\sum_{i=1}^{1}\hat{\varepsilon}_i \\ \vdots \\ -\sum_{i=1}^{T-1}\hat{\varepsilon}_i \end{bmatrix} \\
&= \frac{1}{T^2}\left(\left(\sum_{i=1}^{1}\hat{\varepsilon}_i\right)^2 + \cdots + \left(\sum_{i=1}^{T-1}\hat{\varepsilon}_i\right)^2 + 0 + \left(-\sum_{i=1}^{1}\hat{\varepsilon}_i\right)^2 + \cdots + \left(-\sum_{i=1}^{T-1}\hat{\varepsilon}_i\right)^2\right) \\
&= \frac{2}{T^2}\sum_{s=1}^{T-1}\left(\sum_{i=1}^{s}\hat{\varepsilon}_i\right)^2 \\
&= \frac{2}{T}\sum_{s=1}^{T-1}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{s}\hat{\varepsilon}_i\right)^2
\end{aligned}
$$

We know that $\hat{\varepsilon}_t = Y_t - \bar{Y} = \beta + \varepsilon_t - (\beta + \bar{\varepsilon}) = \varepsilon_t - \bar{\varepsilon}$.

$$
\begin{aligned}
\hat{V}_{KVB} &= \frac{2}{T}\sum_{s=1}^{T-1}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{s}\hat{\varepsilon}_i\right)^2 \\
&= \frac{2}{T}\sum_{s=1}^{T-1}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{s}(\varepsilon_i - \bar{\varepsilon})\right)^2 \\
&= \frac{2}{T}\sum_{s=1}^{T-1}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{s}\varepsilon_i - \frac{s}{\sqrt{T}}\bar{\varepsilon}\right)^2 \\
&= \frac{2}{T}\sum_{s=1}^{T-1}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{s}\varepsilon_i - \frac{s}{T}\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\varepsilon_i\right)^2
\end{aligned}
$$

9.2 Functional CLT

We first introduce the concept of Brownian motion (or the Wiener process).

Definition 9.2.1: Brownian motion

The standard Brownian motion $W(\lambda), \lambda \in [0,1]$ is a continuous time stochastic process such that $W(\lambda_1), \ldots, W(\lambda_k)$ are jointly normally distributed for any $k \in [0,1]$ for fixed $\lambda_1, \ldots, \lambda_k$ with:

$$\mathbb{E}W(\lambda_i) = 0, \quad Cov(W(\lambda_i), W(\lambda_j)) = \min(\lambda_i, \lambda_j) \quad \forall i, j \in [0,1]$$

This is to say, it is a set of random variables indexed by λ, or alternatively a random function in C[0,1] the space of continuous functions on [0,1]. Further any interval of indices within $W(\lambda)$ is jointly normal.

Properties

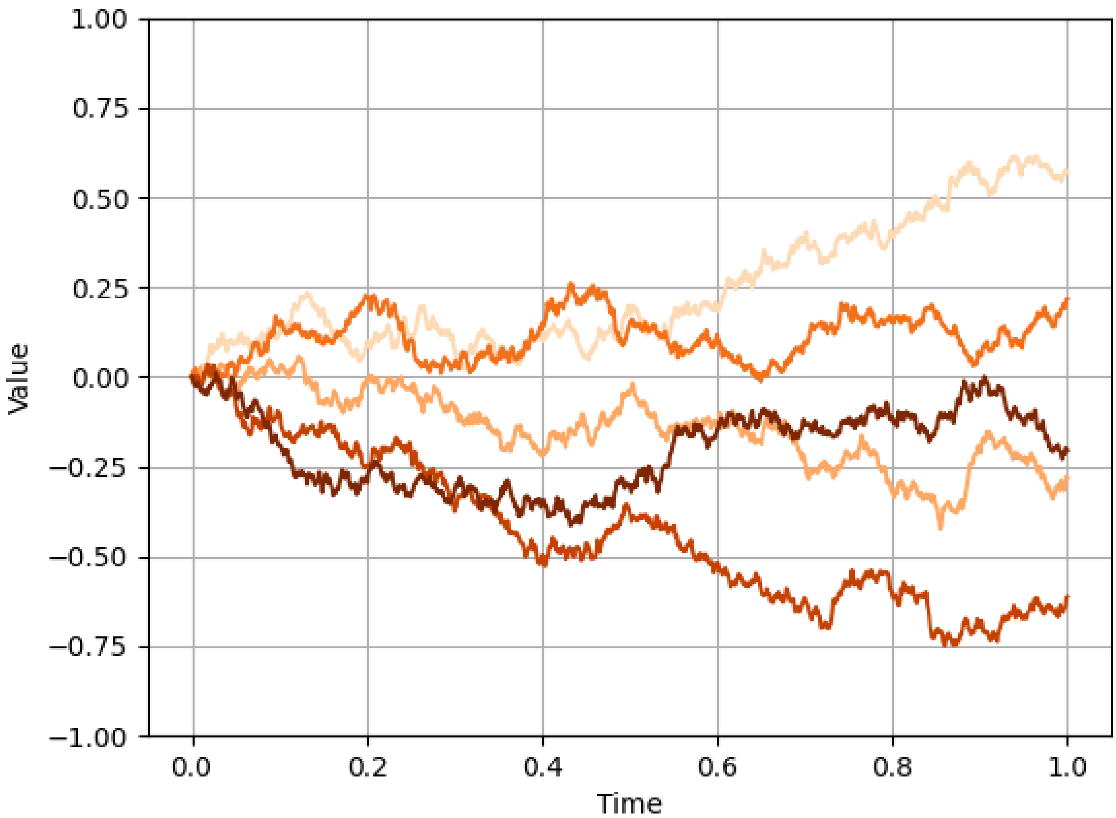
- $\mathrm{Var}(W(\lambda_i)) = \lambda_i$

  $$\mathrm{Var}(W(\lambda_i)) = \mathrm{Cov}(W(\lambda_i), W(\lambda_i)) = \min(\lambda_i, \lambda_i) = \lambda_i$$

- $W(0) = 0$

  $$\mathbb{E}W(0) = 0 \text{ and } \mathrm{Var}(W(0)) = 0$$

- $W(\lambda)$ has independent increments, for every $0 \leq \lambda_1 < \lambda_2 < \cdots < \lambda_k \leq 1$ the random variables $W(\lambda_1), W(\lambda_2) - W(\lambda_1), \ldots, W(\lambda_k) - W(\lambda_{k-1})$ are independent.
- $W(\lambda)$ has gaussian increments, $W(\lambda_{i+u}) - W(\lambda_i) \sim N(0, u)$
- $W(\lambda)$ is nowhere differentiable



The functional central limit theorem (FCLT) is a generalisation of the conventional CLT to function-valued random variables. To understand this we first generalise the standard notions of consistency and convergence in distribution to the space $C[0,1]$. We define the distance between two functions using the sup-norm:

$$d(f,g) = \sup_{x \in [0,1]} |f(x) - g(x)|$$

This represents the maximum distance between the two functions.

Definition 9.2.2: Convergence in probability

A random element $\xi_T \in C[0,1]$ converges in probability to f (that i, $\xi_T \xrightarrow{p} f$) if $\Pr[d(\xi_T, f) > \delta] \to 0$ for all $\delta > 0$.

Definition 9.2.3: Convergence in distribution

Let $\{\xi_T\}$ be a sequence of random elements in $C[0,1]$ and let F be a distribution function on $C[0,1]$, with induced probability measure π_T. Then π_T converges weakly to π, or equivalently $\xi_T \xrightarrow{d} \xi$ where ξ has probability measure π, if and only if $\int f d\pi_T \to f d\pi$ for all bounded continuous functions f: $C[0,1] \to \mathbb{R}$.

Definition 9.2.4: Continuous Mapping Theorem

If h is a continuous functional mapping $C[0,1]$ to some metric space and $\xi_T \xrightarrow{d} \xi$ then $h(\xi_T) \xrightarrow{d} h(\xi)$.

We now present some background and intuition for the functional central limit theorem.

Explanation. Let's first consider the partial sum process, defined as $X_T(\lambda) = \frac{1}{T}\sum_{t=1}^{[T\lambda]} \zeta_t$ with $\zeta \sim WN(0,1)$. The square brackets denote the floor function (i.e. the integer part of $T\lambda$). Let's see how this partial sum looks when $T = 10$ and consider $\xi_T(\lambda)$ for $\lambda = 0, 0.01, 0.1, 0.2$:

$$\lambda = 0, \quad [10 \times 0] = 0: \quad X_{10}(0) = \frac{1}{10}\sum_{t=1}^{0} \zeta_t = 0$$

$$\lambda = 0.01, \quad [10 \times 0.01] = 0: \quad X_{10}(0.01) = \frac{1}{10}\sum_{t=1}^{0} \zeta_t = 0$$

$$\lambda = 0.1, \quad [10 \times 0.1] = 1: \quad X_{10}(0.1) = \frac{1}{10}\sum_{t=1}^{1} \zeta_t = \frac{\zeta_1}{10}$$

$$\lambda = 0.2, \quad [10 \times 0.2] = 2: \quad X_{10}(0.2) = \frac{1}{10}\sum_{t=1}^{2} \zeta_t = \frac{\zeta_1 + \zeta_2}{10}$$

For a sequence of errors ζ_t

1. The function $X_T(\lambda)$ is a random step function defined on [0,1].
2. As T gets bigger the step size gets smaller, and the function becomes smoother (looking more and more like a Wiener process).

Lets consider the following for any fixed $\lambda \in [0,1]$:

$$\begin{aligned}
\sqrt{T} X_T(\lambda) &= \sqrt{T}\frac{1}{T}\sum_{t=1}^{[T\lambda]} \zeta_t \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{[T\lambda]} \zeta_t \\
&= \frac{\sqrt{[T\lambda]}}{\sqrt{T}} \frac{1}{\sqrt{[T\lambda]}}\sum_{t=1}^{[T\lambda]} \zeta_t
\end{aligned}$$

Now, as $T \to \infty$:

$$\frac{\sqrt{[T\lambda]}}{\sqrt{T}} \to \sqrt{\lambda}$$

$$\frac{1}{\sqrt{[T\lambda]}} \sum_{t=1}^{[T\lambda]} \zeta_t \xrightarrow{d} N(0,1) \quad \text{by the CLT}$$

It follows from Slutsky's theorem that:

$$\sqrt{T} X_T(\lambda) \xrightarrow{d} \sqrt{\lambda} N(0,\lambda) = W(\lambda)$$

Since the above holds for any $\lambda \in [0,1]$, we might expect this holds uniformly for $\lambda \in [0,1]$. This is indeed the case, and is known as the functional central limit theorem (or Donsker's theorem for partial sums). □

The above is based on a step-function, however Alexei present a piecewise linear function where we linearly interpolate between points. This is presented below, the substantive results are the same. Let ζ_T, $t = 1, 2, ...$ be zero mean i.i.d. random variables with variance 1. Let $\xi_T(\lambda)$ be the function constructed by linearly interpolating between the partial sums of ζ at the points $\lambda = (0, \frac{1}{T}, \frac{2}{T}, \dots, \frac{T-1}{T}, 1)$, that is:

$$\xi_T(\lambda) = \frac{1}{\sqrt{T}} \left(\sum_{t=1}^{[T\lambda]} \zeta + (T\lambda - [T\lambda]) \zeta_{[T\lambda]+1} \right)$$

so that ξ_T is a piecewise-linear random element of C[0,1] (between each point we linearly interpolate). The CLT for vector valued processes ensures that $[\xi_T(\lambda_1), \xi_T(\lambda_2), \dots, \xi_T(\lambda_k)]$ converges in distribution to a k-dimensional normal random variable. The FCLT extends this result to hold not just for finitely many fixed values of λ, but rather for ξ_T treated as a function of λ.

Theorem 9.2.1 (Functional Central Limit Theorem)**.** $\xi_T(\lambda) \xrightarrow{d} W$, where W is a standard Brownian motion on the unit interval.

Lemma 9.2.1 (Beveridge-Nelson decomposition)**.** Let $u_t \sim I(1)$, where $\Delta u_t = \varepsilon_t = C(L)\zeta_t$. Then

$$u_t = C(1) \sum_{s=1}^{t} \zeta_s + C^*(L)\zeta_t + (u_0 - C^*(L)\zeta_0)$$

Proof.

$$C(L) = C(1) + [C(L) - C(1)] = C(1) + C^*(L)(1-L)$$

where $c_j^* = -\sum_{i=j+1}^{\infty} c_i$. Why can we do this? Define A(L) = C(L) - C(1). Clearly A(1) = 0 and is thus a root of A(L). This justifies the factorisation C(L)-C(1) = (1-L)C*(L).
Thus we can write

$$\varepsilon_t = C(L)\zeta_t = C(1)\zeta_t + C^*(L)\Delta\zeta_t$$

Then because $u_t = \sum_{s=1}^t \varepsilon_s + u_0$ we get the result:

$$
\begin{aligned}
u_t = \sum_{s=1}^{t} \varepsilon_s + u_0 &= \sum_{s=1}^{t} C(1)\zeta_s + C^*(L)\Delta\zeta_s + u_0 \\
&= C(1)\sum_{s=1}^{t}\zeta_s + C^*(L)\zeta_t + u_0 - C^*(L)\zeta_0
\end{aligned}
$$

□

We now consider some arbitrary linear process $\varepsilon_t = C(L)\zeta_t$ where $\zeta_t \sim iid(0,1)$ as before. By the B-N decomposition we can write $\varepsilon_t = C(1)\zeta_t + C^*(L)\Delta\zeta_t$. Consider the following:

$$
\begin{aligned}
\nu_T(\lambda) &= \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{[T\lambda]}\varepsilon_t + (T\lambda - [T\lambda])\varepsilon_{[T\lambda]+1}\right) \\
&= \frac{1}{\sqrt{T}}\left(\sum_{t=1}^{[T\lambda]}(C(1)\zeta_t + C^*(L)\Delta\zeta_t) + (T\lambda - [T\lambda])C(1)(\zeta_{[T\lambda]+1} + C^*(L)\Delta\zeta_{[T\lambda]+1})\right) \\
&= \frac{1}{\sqrt{T}}\left(C(1)\sum_{t=1}^{[T\lambda]}\zeta_t + C^*(L)\zeta_{[T\lambda]} - C^*(L)\zeta_0 + (T\lambda - [T\lambda])C(1)(\zeta_{[T\lambda]+1} + C^*(L)\Delta\zeta_{[T\lambda]+1})\right) \\
&= C(1)\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{[T\lambda]}\zeta_t + (T\lambda - [T\lambda])\zeta_{[T\lambda]+1}\right) + C^*(L)\frac{1}{\sqrt{T}}\left(\zeta_{[T\lambda]} - \zeta_0 + (T\lambda - [T\lambda])\Delta\zeta_{[T\lambda]+1}\right) \\
&= C(1)\xi_T(\lambda) + \frac{1}{\sqrt{T}}I(0)
\end{aligned}
$$

Since the second term is $T^{-\frac{1}{2}}$ multiplied by an I(0) process, it converges to zero in probability. Further, since $\xi_T(\lambda) \xrightarrow{d} W(\lambda)$, this suggests that $C(1)\xi_T \xrightarrow{d} C(1)W(\lambda)$ and $\nu_T(\lambda) \xrightarrow{d} C(1)W(\lambda)$.

Theorem 9.2.2. $\nu_T(\lambda) \xrightarrow{d} C(1)W(\lambda)$

For a more rigorous proof of convergence see Stock (1994) pg 2750. [1]

9.3 Fixed-b asymptotics

Recall the definition of Brownian motion (ignoring the smoothing terms):

$$
\xi_T(\lambda) = \frac{1}{\sqrt{T}}\sum_{t=1}^{[T\lambda]}\varepsilon_t.
$$

Thus we can see that

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{i=1}^{s}\varepsilon_i = \frac{1}{\sqrt{T}}\sum_{i=1}^{T\times\frac{s}{T}}\varepsilon_i = \xi_T(\frac{s}{T}) \\
&\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\varepsilon_i = \xi_T(1)
\end{aligned}
$$

[1]This topic is such a fucking rabbit hole, there is no chance this is understandable to our tiny reg monkey brains. This shit is so convoluted don't even bother going further.

Consider our representation from earlier:

$$\begin{aligned}
\hat{V}_{KVB} &= \frac{2}{T}\sum_{s=1}^{T-1}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{s}\varepsilon_i - \frac{s}{T}\frac{1}{\sqrt{T}}\sum_{i=1}^{T}\varepsilon_i\right)^2 \\
&= \frac{2}{T}\sum_{s=1}^{T-1}\left(\xi_T(\frac{s}{T}) - \frac{s}{T}\xi_T(1)\right)^2 \\
&\approx 2\int_0^1 \left(\xi_T(\lambda) - \lambda\xi_T(1)\right)^2 d\lambda \quad \lambda := \frac{s}{T}
\end{aligned}$$

The approximation follows from the fact that the second line is a Riemann sum, where as $T \to \infty$ the approximation error converges to zero.
We know that $\xi_T(\lambda) \xrightarrow{d} c(1)W(\lambda)$, thus by the continuous mapping theorem:

$$\begin{aligned}
\hat{V}_{KVB} &\xrightarrow{d} 2\int_0^1 \left(c(1)W(\lambda) - \lambda c(1)W(1)\right)^2 d\lambda \\
&= 2[c(1)]^2\int_0^1 \left(W(\lambda) - \lambda W(1)\right)^2 d\lambda
\end{aligned}$$

The right hand side is proportional to $[c(1)]^2$, which is the long-run variance of ε_t.

Example (Long-run variance). $\varepsilon_t = C(L)\zeta_t = c_0\zeta_t + c_1\zeta_{t-1} + \ldots$
Long run variance is defined differently to before, here it is $\mathrm{Var}(\varepsilon_t)$.

$$\begin{aligned}
Var(\varepsilon_t) &= Var(C(L)\zeta_t) \\
&= Var(C(1)\zeta_t) \quad \text{since } \zeta_t \text{ is i.i.d. the lags don't matter} \\
&= C(1)^2 Var(\zeta_t) \\
&= C(1)^2 \quad \text{since } \zeta_t \sim iid(0,1)
\end{aligned}$$

If we now consider the t-statistic (based on $\hat{V}_{KVB}$) for testing $H_0 : \beta = 0$:

$$\begin{aligned}
t &= \frac{\hat{\beta}}{\sqrt{Var(\hat{\beta})}} = \frac{\bar{Y}}{\sqrt{\frac{1}{T}\hat{V}_{KVB}}} = \frac{\beta + \bar{\varepsilon}}{\frac{1}{\sqrt{T}}\sqrt{\hat{V}_{KVB}}} \\
&\overset{H_0}{=} \frac{\sqrt{T}\bar{\varepsilon}}{\sqrt{\hat{V}_{KVB}}} = \frac{\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\varepsilon_j}{\sqrt{\hat{V}_{KVB}}} = \frac{\frac{1}{\sqrt{T}}\xi_T(1)}{\sqrt{\hat{V}_{KVB}}} \\
&\xrightarrow{d} \frac{c(1)W(1)}{\sqrt{2[c(1)]^2\int_0^1 \left(W(\lambda) - \lambda W(1)\right)^2 d\lambda}} \\
&= \frac{c(1)W(1)}{c(1)\sqrt{2\int_0^1 \left(W(\lambda) - \lambda W(1)\right)^2 d\lambda}} \\
&= \frac{W(1)}{\sqrt{2\int_0^1 \left(W(\lambda) - \lambda W(1)\right)^2 d\lambda}}
\end{aligned}$$

This doesn't depend on $c(1)$ (the model parameters), meaning the distribution is pivotal. Thus it can be simulated and critical values recorded. The pdf is given below, note how the (normalised) KVB distribution has fatter tails than the normal distribution.

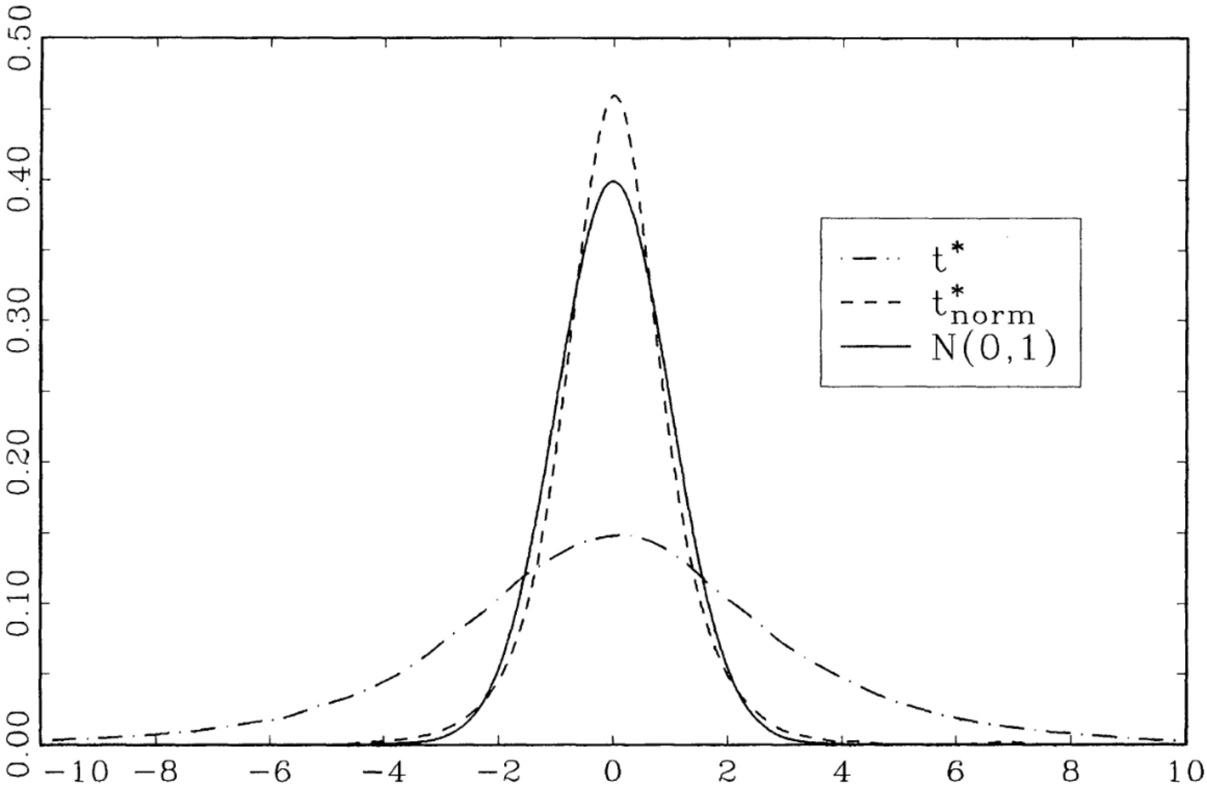


FIGURE 1.—Densities of t^*, t^*_{norm}, and $N(0, 1)$.

KVB show that in finite samples these tests may outperform tests based on Newey-West standard errors. At a high level, if there is lots of serial correlation KVB is much better, whereas if it is only minor NW is probably fine. NW suffers when samples are small and serial correlation is large. KVB, like HAC estimator tests, suffer from serious size distortions (although less so) if the data have highly persistent serial correlation and are close to being non-stationary. KVB also show the finite sample power of their test dominates finite sample power of HAC tests.