

Joakim, Westerlund; Johan, Blomquist

Working Paper

Are Crime Rates Really Stationary?

Working Paper, No. 2009:20

Provided in Cooperation with:

Department of Economics, School of Economics and Management, Lund University

Suggested Citation: Joakim, Westerlund; Johan, Blomquist (2009) : Are Crime Rates Really Stationary?, Working Paper, No. 2009:20, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/259979>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

ARE CRIME RATES REALLY STATIONARY?*

Joakim Westerlund[†]
University of Gothenburg
Sweden

Johan Blomquist
Lund University
Sweden

December 21, 2009

Abstract

Many empirical studies of the economics of crime focus solely on the determinants thereof, and do not consider the dynamic and cross-sectional properties of their data. As a response to this, the current paper offers an in-depth analysis of this issue using data covering 21 Swedish counties from 1975 to 2008. The results suggest that the four crime types considered are non-stationary, and that this cannot be attributed to county specific disparities, but rather that it is due to a small number of common stochastic trends to which groups of counties tend to revert. The results further suggest that these trends can be given a macroeconomic interpretation. Our findings are consistent with recent theoretical models predicting that crime should be dependent across both time and counties.

JEL Classification: C32; C33; E20; K40.

Keywords: Crime; Non-stationary data; Panel unit root tests; Common factor.

1 Introduction

Crime rates usually exhibit substantial variation across time. Indeed, the total number of offences recorded by the Swedish police per 100,000 of the population has gone from 9,223 in 1975 to 14,112 in 2007, an increase by more than 50%. But there is not only the time

*Previous versions of this paper were presented at the 5th Nordic Econometric Meeting in Lund and at a seminar at Lund University. The authors would like to thank seminar participants and in particular David Edgerton, Randi Hjalmarsson, Matthew Lindquist, Peter Lindström, Hashem Pesaran and three anonymous referees for may valuable comments and suggestions. The authors gratefully acknowledge financial support from the Jan Wallander and Tom Hedelius Foundation, research grant numbers W2006–0068:1 and P2009–0189:1.

[†]Corresponding author: Department of Economics, University of Gothenburg, P. O. Box 640, SE-405 30 Gothenburg, Sweden. Telephone: +46 31 786 5251, Fax: +46 31 786 1043, E-mail address: joakim.westerlund@economics.gu.se.

series variation, there is also the cross-sectional variation, which is just as pronounced. For example, in 2001 the number of thefts and robberies per capita reported in the capital of Stockholm was 0.09, which is almost two times as many as in the rural southern county of Blekinge. The most northern county of Norrbotten has a similar, low, crime rate of 0.05, whereas in Skåne, which is a neighboring county of Blekinge, the crime rate is almost as high as in Stockholm.

A common explanation for this variation is that it is due to differing economic conditions. But these differences are usually not nearly enough to account for the full extent of the cross-sectional variation. For example, in 2001 the unemployment rate in Stockholm was 2.68%, which is low when compared to 4.44% in Blekinge and 5.26% in Skåne. The relatively high crime rates in Stockholm and Skåne also coexisted with much higher income levels when compared to Blekinge and Norrbotten.¹

As a response to this, a new class of models that stresses the importance of social interactions has emerged, see for example Sah (1991) and Glaeser *et al.* (1996). The main lesson being that static models are not enough to capture the behavior of crime. These models therefore predict that crime should be persistent over time, and some are even admitting to the possibility that crime may be non-stationary.

1.1 Limitations of earlier studies

Although theory tells us that crime should be persistent, this lesson is only rarely taken into consideration when conducting empirical work. In fact, even the most recent research tend to focus on static regressions, which is problematic for at least two reasons.

Firstly, the dynamics of crime can have implications for policy that are neglected when using static regressions. Suppose for example that there is a temporary policy shock in the rate of unemployment that rises the number of crimes committed. If crime is persistent then this shock will be carried forward into the future. By using static regressions we ignore this possibility, which may well lead to a misstatement of the effect of policy actions on current and future crime rates.

Secondly, the presence of unattended dynamics may compromise inference, and in the extreme case when crime is non-stationary inference may even be spurious. Take for exam-

¹The fact that only a small fraction of the cross-regional crime variation can be explained by differences in economic conditions can also be observed in data for the United States, see Glaeser *et al.* (1996).

ple the study of Edmark (2005) who uses Swedish county-level data between 1988 and 1999 to estimate the relationship between unemployment and property crime. Although many of her panel regressions have R^2 statistics that are very close to unity, a well-known sign of spuriousness, the unit root hypothesis is never tested.

Of course, these problems of neglected dynamics are not unique to panel data. But if one admits to the possibility of an heterogeneous data generating process with different dynamics for each unit, then there is not just one potential error to be made but as many as there are units in the panel. The effect of omitted dynamics is therefore likely to increase, and to become even more severe as the cross-sectional dimension of the panel increases.

There are of course studies that do allow for dynamics and even unit roots. But these are almost exclusively based on aggregated time series data, usually at the country level, which means that the cross-sectional variation is effectively ignored.² Similarly, while there have been attempts to allow for dynamics in panels of disaggregated crime data, in these studies there is usually no room for any interactions between the panel members, which is just as problematic as when ignoring the dynamics.³ In fact, most theoretical models predicting that crime should be persistent also predict that there should be at least some form of interaction across the cross-sectional units, see Sah (1991).⁴

Other studies use static regressions that are augmented with a linear time trend to account for the fact that crime is usually trending, see for example Gould *et al.* (2002) and Raphael and Winter-Ebmer (2001), who document a positive relationship between unemployment and property crime. Apart from the cross-sectional independence assumption, which is almost always there, the main problem here is that the trend is assumed to be de-

²There is also the issue of aggregation. When studying time series properties of crime, researchers typically rely on national crime rate data, see for example Hale (1998), Fajnzylber *et al.* (2002), and Funk and Kugler (2003). The question is of how well the behavior of crime at the aggregate country level approximates the properties of the individual regions? This question cannot be examined when only aggregate data are available. However, when data are available at a disaggregate level, it is quite well known that the regional features may not be preserved at the country level, and a crucial role is played by the degree of heterogeneity amongst the regions. In fact, as Forni and Lippi (1997) show, irrespective of the approach one chooses for the aggregate analysis, when heterogeneity is allowed, the dynamic properties of the aggregated data differ from those of the disaggregated data, thereby leading to substantially different interpretations.

³Take as an example the study of Fajnzylber *et al.* (2002), in which a dynamic panel regression is fitted to country-level data. Although the results indicate that there is a link between violent crimes and economic growth and income inequality, since the countries are assumed to be independent, there is no way of knowing whether this link represents a true casual relationship or if it is just an artifact of omitted cross-country interdependencies.

⁴Using city-level data for the United States between 1960 and 2004, McDowall and Loftin (2009) find that nationwide factors accounted for more than 20% of the total variation in their data.

terministic. In other words, while recognizing the presence of a trend, these studies do not allow for the possibility that it might be stochastic.

1.2 Recent developments and the main results of this study

As the above discussion makes clear, while reasonable and potentially appealing, most of the earlier empirical approaches have been inadequate and not very convincing, and this paper therefore proposes an alternative approach. The idea is that to be able to provide any reliable evidence on the behavior of crime one needs to consider not the time series and cross-sectional variation separately but simultaneously.

This idea is not completely new, of course. The first attempt to combine the two sources of variation that we can find appears in Witt *et al.* (1998). The motivation for their paper is that if regional crime rates are non-stationary, then there is also a possibility that they might be cointegrated with each other, a situation very much related to what is commonly referred to in the growth literature as club convergence.⁵ That is, although individually diverging, there might still be clubs of regions that are converging along a common stochastic trend. Using data that cover four English regions between 1975 and 1996, the authors find evidence of such a trend, suggesting the existence of a unique long-run relationship between the four regions.

The problem is that the econometric approach is a multivariate one, which cannot handle panels unless the cross-sectional dimension N very small. In fact, for this approach to work properly, not only must N be small enough, the time series dimension T has to be substantial, a condition that is rarely fulfilled in practice. Thus, what is really needed here is a panel approach that is applicable even in situations when N is large, and the current paper can be seen as an attempt in this direction. Another problem with the Witt *et al.* (1998) study is that it does not provide any insight as to what the common stochastic trend actually represents. Is it for example due to common business cycle variations or it is maybe due some policy shock?

Our starting point is the panel analysis of non-stationary idiosyncratic and common components, or PANIC, method of Bai and Ng (2004). The idea is to first decompose the observed data into two components, one that is common to all regions and one that is idiosyncratic or

⁵Actually, cross-region cointegration is just a necessary condition for club convergence, see Section 4.3 for a discussion.

region-specific. The objective of PANIC is then to infer the order of integration of the data by testing for unit roots in each component separately. The main advantage of this approach in comparison to the one used by Witt *et al.* (1998) is that here N does not have to be small.

There are of course other panel unit root tests around that allow for cross-section dependence in a large N environment, see Breitung and Pesaran (2008) for a recent survey of the literature. However, PANIC has a number of distinct features that makes it more suitable for our purposes. The main advantage is that in contrast to most, if not all, other tests around, in PANIC the common and idiosyncratic components are not restricted to have the same order of integration. This is an important merit in our application since previous research emphasizes that regional crime rates are influenced both by nationwide and region-specific factors, see for example McDowall and Loftin (2009). The PANIC procedure allows us to test for unit roots in both components separately without any *a priori* knowledge about the integratedness of the data. Another advantage is that in PANIC the cross-section dependence is not treated as nuisance, but rather as an object of interest. This is especially valuable in cases such as this when the forces driving the common variation could provide useful information for policy. PANIC is also very simple to implement and it has been shown to perform well in samples as small as ours, see for example Gutierrez (2006).

One problem with PANIC is that it is not equipped to handle cases when there is uncertainty over the presence of the deterministic time trend, which is of course always the case in practice. Therefore, in order to account for this uncertainty, a sequential test procedure is proposed to determine the extent of both the trend and the non-stationarity of the panel.

The data that we use cover 21 Swedish counties between 1975 and 2008, which means that there are 714 observations available for each of the four crimes considered, burglary, theft, robbery and homicide. The results suggest that all four crimes are non-stationary, and that this cannot be attributed to county specific disparities but rather that it is due to the presence of common stochastic trends. We also find that these stochastic trends are driven in part by macroeconomic conditions.

The rest of the paper is organized as follows. In section 2 we briefly discuss the theoretical motivation for the PANIC approach. Section 3 describes the methodology in detail. Section 4 then presents the data that are used and reports the results of the empirical analysis. Section 5 concludes.

2 Theoretical considerations

In his seminal paper, Becker (1968) develops the basic elements of a deterrence model, wherein the choice of the individual of whether to engage in crime or not is viewed as a function of the relationship between the expected benefits of crime and the expected costs of punishment. The model predicts how exogenous changes in the probability of arrest may affect expected payoffs, and thus the supply of crime. The logic is disarmingly simple. By setting the expected cost high enough to dissuade the individual from choosing to commit illegal acts, deterrence should reduce crime.

But as Sah (1991) points out in general things are not so simple. He questions the preference of Becker (1968) to treat the expected cost of punishment as exogenous, which implies that crime should be completely static. Sah (1991) therefore endogenizes the expected cost of punishment, and in the process of doing so, he develops a model in which the expected cost, and therefore also crime itself, is time-varying. The basic idea is that if the probability of arrest has to be learned from others or through experience, the effects of policy will take time to materialize. Murphy *et al.* (1993) and Glaeser *et al.* (1996) further show how social interactions between criminals can make the effects of policy even more persistent.

A key insight from these models is therefore that crime should be persistent, with temporary shocks having long-lasting effects. The extreme case is that of a unit root, which would entail permanent effects. But this is not the only insight. Indeed, as Sah (1991) points out, if counties are highly segregated, then county-level crime rates are expected to differ significantly, whereas if there is some intercounty interaction, then criminality is expected to spill over across counties. A common rationale for this interaction among counties is that it represents factors such as common business cycle fluctuations, demographic composition, and common crime-fighting policies, see Hale (1998), Levitt (1999), and Funk and Kugler (2003), among others. In other words, crime should not only be dependent across time but potentially also across counties.

3 Econometric methodology

3.1 PANIC

Consider the crime rate X_{it} , observable for $i = 1, \dots, N$ counties and $t = 1, \dots, T$ years. The preceding discussion suggests that the possible non-stationarity of X_{it} can originate from

two sources, one that is common across counties and one that is idiosyncratic. This is also the starting point of PANIC. Let us therefore decompose X_{it} as

$$X_{it} = D_{it} + \lambda_i' F_t + e_{it}, \quad (1)$$

where D_{it} represents the deterministic component of crime, whose specification is going to turn out to be very important later on. Typically D_{it} is just an intercept but in this paper we set

$$D_{it} = c_i + \beta_i t, \quad (2)$$

thereby admitting to the possibility that X_{it} might be trending deterministically.

The common factor F_t and loading λ_i together represent the common component of X_{it} , where the j^{th} element of the r -dimensional vector F_t is assumed to follow a first-order autoregressive process with a possibly serially correlated error term η_{jt} ,

$$F_{jt} = \alpha_j F_{jt-1} + \eta_{jt}. \quad (3)$$

As for the county-specific, or idiosyncratic, component e_{it} , we make a similar assumption. In particular, we assume that

$$e_{it} = \delta_i e_{it-1} + \epsilon_{it}, \quad (4)$$

where ϵ_{it} may be correlated across t but not across i , which is the same as saying that any dependence is captured by the common component.⁶ The type of serial correlation that can be permitted is very general, and include for example the broad class of stationary autoregressive moving average models.⁷ The errors ϵ_{it} and η_{jt} , which are mutually independent for all i, j and t , are homoskedastic in t but not necessarily in i and j .

As a response to the poor precision and power of conventional time series unit root tests, Moody and Marvell (2005), and Phillips (2006) apply a battery of so-called first generation panel unit root tests, with which they are able to reject the presence of unit roots in crime rates for the United States. Unfortunately, these tests are only appropriate if the states are

⁶The assumption that ϵ_{it} is cross-sectionally independent is actually only for simplicity and can be relaxed at the cost of some extra moment conditions to allow for weak cross-sectional dependence, see Appendix A for a detailed discussion.

⁷In fact, the only restriction is that the partial sum processes of ϵ_{it} and η_{jt} satisfy a functional central limit theorem, see Appendix A.

uncorrelated, and hence cannot be used for analyzing more complex issues of interstate dependency, such as cross-sectional cointegration. In terms of the model in (1) to (4), the first generation tests assume that there is no common component, and hence that X_{it} is completely idiosyncratic. It also implies that e_{it} is the only source of potential non-stationarity. Our model is more general, and allows for cross-county dependence, as well as an additional source of stationarity, F_t . Thus, in this model, the possible non-stationarity of X_{it} can originate from F_t or e_{it} , or both. We also allow the autoregressive behavior to differ across both factors and counties, so that for example some of the factors may be non-stationary while other are not.

Whether these components actually are stationary or not is an empirical matter. The problem is that F_t and e_{it} are unobserved, which of course makes all forms of unit root testing impossible. The first step in PANIC is therefore to try to estimate these components, which can be done by using the method of principal components. However, since in this paper crime may be non-stationary this method cannot be applied to the level data, as this might result in factor estimates that are non-stationary even though the true factors are stationary. We therefore consider the first-differenced data,

$$\Delta X_{it} = \Delta D_{it} + \lambda'_i \Delta F_t + \Delta e_{it}, \quad (5)$$

which are mean zero and stationary as long as D_{it} does not contain a trend.⁸ To eliminate the nonzero mean in case of a trend we further demean ΔX_{it} , giving

$$\Delta X_{it} - \overline{\Delta X}_i = \lambda'_i (\Delta F_t - \overline{\Delta F}) + \Delta e_{it} - \overline{\Delta e}_i, \quad (6)$$

where $\overline{\Delta X}_i = \frac{1}{T-1} \sum_{t=2}^T \Delta X_{it}$ with an obvious definition of $\overline{\Delta F}$ and $\overline{\Delta e}_i$. By applying the principal components method to either ΔX_{it} or $\Delta X_{it} - \overline{\Delta X}_i$ we obtain estimates of the components in first differences, denoted $\Delta \hat{F}_t$ and $\Delta \hat{e}_{it}$, which can then be accumulated to obtain the corresponding level estimates, henceforth denoted \hat{F}_t and \hat{e}_{it} , respectively.

Having obtained \hat{F}_t and \hat{e}_{it} , PANIC then proceeds to test the two components for unit roots, thereby making it possible to disentangle the sources of potential non-stationarity in X_{it} . If the non-stationary is due to \hat{F}_t , then X_{it} is diverging along a common stochastic trend, while if the non-stationary is due to \hat{e}_{it} , then the divergence is due to county specific sources. If \hat{F}_t is non-stationary, while if \hat{e}_{it} is stationary, crime is cointegrated across counties,

⁸Note that since F_t and e_{it} are assumed to be integrated of at most order one, ΔX_{it} must be stationary.

permitting for the possibility of different convergence clubs. Finally, if \hat{F}_t and \hat{e}_{it} are both non-stationary, then the divergence has two sources, one that is common and one that is idiosyncratic.

The justification for testing in this particular way is that the unit root test of \hat{e}_{it} is asymptotically equivalent to that of e_{it} . Similarly, knowing \hat{F}_t is as good as knowing HF_t , in the sense that testing \hat{F}_t is asymptotically equivalent to testing HF_t , where H is an $r \times r$ rotation matrix of full rank that accounts for the fact that λ_i and F_t are not separately identifiable.⁹

One implication of this is that since \hat{e}_{it} is asymptotically independent of \hat{F}_t , there is no need for a joint test, which of course makes the testing very simple. Moreover, because \hat{e}_{it} is consistent for e_{it} , which in turn is independent across i , the testing of \hat{e}_{it} can be conducted by using any conventional first generation panel unit root test, and so there is no need for a special test. Bai and Ng (2004) recommend using the meta approach of Choi (2001), which is based on combining the p -values from the well-known augmented Dickey and Fuller (1979) test, henceforth denoted ADF, when applied to each county. The resulting panel test, henceforth denoted P , has been shown to work very well, even in small samples such as ours, and will therefore be used also in this paper. For testing the common component, Bai and Ng (2004) propose using the ADF test.

3.2 Testing for the presence of a trend

Although very general when it comes to the allowable forms of serial and cross-sectional correlation, the standard PANIC procedure as proposed by Bai and Ng (2004) is still rather restrictive in the sense that it assumes that the researcher knows with full certainty whether or not the trend should be included in D_{it} , which is of course never the case in practice. This is problematic for at least two reasons.

The first problem is how to deal with this uncertainty in practice. In the time series literature unit root tests are often conducted after at least some form of pre-testing for the trend, taking the constant term as given. Most of the time these pre-tests are rather informal, involving for example inspection of plots of the data and significance tests of the trend slope in the fitted test regression. Regardless of whether such pre-tests are employed or not, it is very common to implement the unit root test both with and without the trend, oftentimes

⁹As is well known, the factor model in (5) is fundamentally unidentified because $\lambda_i'HH^{-1}F_t = \lambda_i'F_t$ for any invertible matrix H . However, in our case exact identification of the true factors F_t is not necessary as the cointegrating rank of F_t is the same as the cointegrating rank of HF_t .

with conflicting results. Indeed, most empirical work tend to suggest that test results can be highly sensitive to the treatment of the trend.

In panels, the decision of whether to include the trend or not is even more complex, especially if one admits to the possibility of unit-specific trend slopes, in which case the choice must be made not just once but N times, at least in principle. The sensitivity to the treatment of the trend is therefore usually much higher in panels than in single time series. In spite of this, researchers that work with panels tend to use much less pre-testing than researchers that work with time series. A common response to the greater uncertainty over the trend component is therefore to simply ignore it.

The second problem is more theoretical in nature and refers to the statistical properties of the PANIC procedure when it is not certain whether the trend should be included. To appreciate the issues involved Table 1 reports some results of the size and power of the ADF and P tests when the significance level is 5%. For simplicity, the data are generated from (1) to (4) with $r = 1$ and $\lambda_i \sim N(1, 1)$ but otherwise equal coefficients for all i . In particular, the deterministic component in (2) is specified with $c_i = 1$ and $\beta_i = \beta$. The errors in (3) and (4) are both drawn from the standard normal distribution.¹⁰

In agreement with the theoretical prediction on this issue we see that both tests are biased towards the null if the regression is fitted with an intercept but the data are generated with both a constant and trend. In other words, the trend can be mistaken for a unit root, which is also the reason for why one cannot run trend augmented regressions without first testing whether the observed trend is truly deterministic, as in for example Gould *et al.* (2002) and Raphael and Winter-Ebmer (2001). On the other hand, if the data are generated with a constant, then we see that the inclusion of a trend leads to a loss of power, which can sometimes be substantial, especially when N and T are small. Only if the deterministic component is specified correctly do the tests enjoy both high power and good size accuracy.

In order to eliminate these adverse effects we look for a procedure that can be used to test for the presence of a trend, and that does not suffer too much from the uncertainty about the integratedness of the data. This is not easy because unlike in the conventional testing situation here we have two potential unit root sources, and so it is not even certain from where any non-stationarity originates. One implication of this is that we have to decide

¹⁰We use 5,000 replications with 100 startup values for each unit throughout. All initial conditions are set to zero.

upon the presence of the trend already before the two components are estimated. If there is no trend then the principal components method is applied to ΔX_{it} , whereas if there is a trend then it is applied to $\Delta X_{it} - \overline{\Delta X}_i$. Moreover, once the trend has been removed it is no longer possible to test for its presence by using the estimated components.

Our solution to this problem is very simple and starts with the regression in (5), which in the case of a trend is given by

$$\Delta X_{it} = \beta_i + \lambda_i' \Delta F_t + \Delta e_{it}.$$

Letting $\overline{\Delta X}_t = \frac{1}{N} \sum_{i=1}^N \Delta X_{it}$ with a similar definition of $\overline{\beta}$, $\overline{\lambda}$ and $\overline{\Delta e}_t$, by averaging across i ,

$$\overline{\Delta X}_t = \overline{\beta} + \overline{\lambda}' \Delta F_t + \overline{\Delta e}_t \quad (7)$$

and by further averaging across t ,

$$\overline{\Delta X} = \overline{\beta} + \overline{\lambda}' \overline{\Delta F} + \overline{\Delta e} \quad (8)$$

with an obvious definition of $\overline{\Delta X}$, $\overline{\Delta F}$ and $\overline{\Delta e}$.

The null hypothesis of interest is that of no trend, that is, $\beta_i = 0$ for all i . Let $\hat{\beta} = \overline{\Delta X}$ denote the first difference estimator of $\overline{\beta}$, and let

$$\hat{\omega}^2 = \frac{1}{T-1} \sum_{t=2}^T (\overline{\Delta X}_t - \hat{\beta})^2 + \frac{2}{T-1} \sum_{s=1}^{M-1} K(s) \sum_{t=s+1}^T (\overline{\Delta X}_t - \hat{\beta})(\overline{\Delta X}_{t-s} - \hat{\beta})$$

be the conventional Newey and West (1994) estimator of the so-called long-run variance of $\overline{\Delta X}_t$, where $K(s) = 1 - \frac{s}{M}$ is the Bartlett kernel and M is the bandwidth parameter that determines how many autocovariances to include in the estimation. In Appendix A we show that as long as the number of unit roots in F_t , henceforth denoted r_1 , is not zero,

$$t_\beta = \frac{\sqrt{T}(\hat{\beta} - \overline{\beta})}{\hat{\omega}} \rightarrow_d N(0, 1) \quad (9)$$

as $M, N, T \rightarrow \infty$ with $M/\min\{N, \sqrt{T}\} \rightarrow 0$. Note in particular that under the null hypothesis of no trend, $t_\beta = \frac{\sqrt{T}\hat{\beta}}{\hat{\omega}} \rightarrow_d N(0, 1)$, suggesting that $|t_\beta|$ can be used to determine whether the trend should be included or not.

The working behind the result in (9) is the following. Note first that since ΔF_t and Δe_{it} are mean zero, $E(\overline{\Delta X}) = \overline{\beta}$, showing that $\overline{\Delta X}$ is an unbiased estimator of $\overline{\beta}$. In fact, since Δe_{it} is independent across i , we even have

$$\overline{\Delta X}_t = \overline{\beta} + \overline{\lambda}' \Delta F_t + o_p(1),$$

suggesting that asymptotically $\overline{\Delta X}_t$ is just a linear combination of the elements of ΔF_t , which is where the asymptotic distribution of t_β comes from.¹¹ This is basically the same trick underlying the proposal of Pesaran (2006, 2007a), which in the current context involves using $\overline{\Delta X}_t$ to approximate the common factors in ΔF_t . A possible alternative to using the method of principal components to estimate the factors in the unit root test is therefore to use the cross-sectional averages. At this point, however, we are only interested in making inference about $\bar{\beta}$, and for this purpose we can make use of the fact that as long as F_t has at least one unit root, then $\overline{\Delta F} = O_p(\sqrt{T})$, implying

$$\sqrt{T}(\overline{\Delta X} - \bar{\beta}) = \bar{\lambda}' \sqrt{T} \overline{\Delta F} + \sqrt{T} \overline{\Delta e} = O_p(1).$$

In Appendix A we show that the $O_p(1)$ term is actually normal, and this is the main insight behind (9).

In order to evaluate the extent to which this asymptotic result applies in small samples we again use simulations. Table 2 reports some results from the size of a double-sided 5% level test when the data are generated as before but now with $r = 5$ and the null of a zero trend slope imposed. To evaluate the effect of serial correlation in the error driving F_{jt} we set $\eta_{jt} = \rho \eta_{jt-1} + v_{jt}$, where $v_{jt} \sim N(0, 1)$. Three different rules for the choice of the bandwidth M are considered. The first is the data dependent rule of Newey and West (1994), while the remaining two rules are deterministic, and involve setting M either equal to $4(T/100)^{2/9}$ as suggested by Newey and West (1994) or equal to zero as when ignoring the effect of serial correlation.

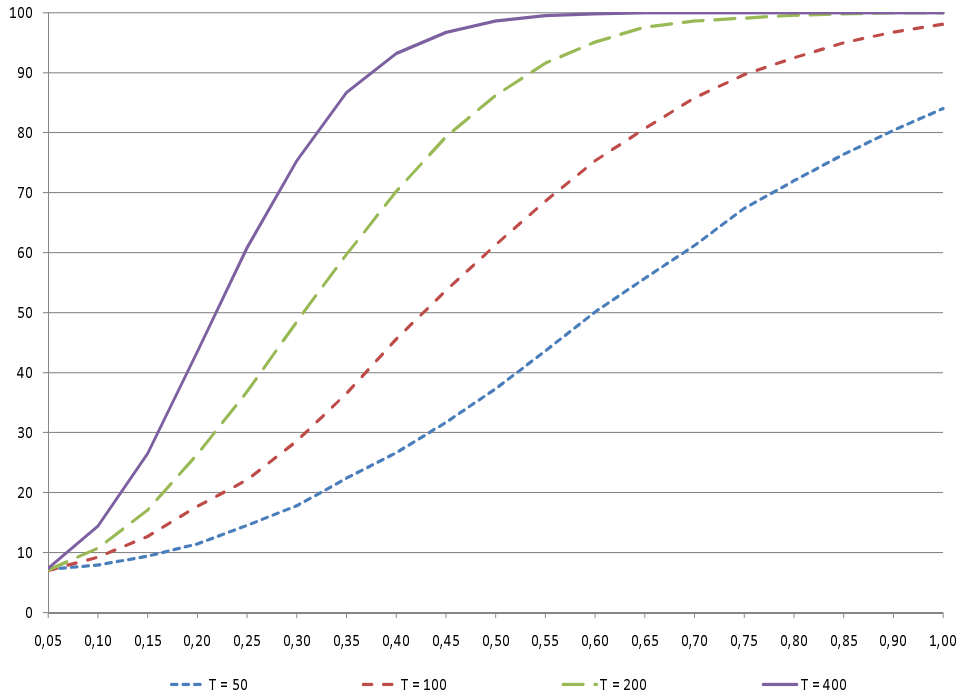
As expected we see that the test performs well when $\rho = 0$ with good size accuracy for all combinations of N and T . The only exception is when $r_1 = 3$, in which case there is a tendency for the test to become undersized. The reason for this lies with the rate of consistency of $\hat{\omega}^2$, which goes from $M/\min\{N, \sqrt{T}\}$ when $r_1 = r$ to $1/M$ when $r_1 < r$, see Appendix A. Thus, as expected given the requirement that $M \rightarrow \infty$, setting M to a small number is not expected to work, and therefore $M = 0$ is not expected to work at all.¹² The fact that the size distortions go in the opposite direction when $\rho = 0.3$ is also

¹¹For any real r , $a_T = O_p(T^r)$ will henceforth be used to indicate that a_T is at most of order T^r in probability, meaning means that a_T/T^r converges in distribution as T grows, while $a_T = o_p(T^r)$ will be used in case a_T is of smaller order in probability than T^r .

¹²The condition that M should go to infinity with the sample size is standard even in the time series literature. The reason for this is the presence of bias, whose elimination requires letting M to infinity, but at a slower rate than \sqrt{T} . The condition that $M/\min\{N, \sqrt{T}\} \rightarrow 0$ as $M, N, T \rightarrow \infty$ can be seen as a panel version of this

partly expected given the well-documented effects of positive serial correlation, and so is their relative magnitude when $M = 0$. The overall best performance is obtained by using the data dependent bandwidth rule.¹³

Figure 1: Power for different values of T .



Next, we consider some results from the power of the test, which are summarized in Figures 1 to 3. In Figure 1, we plot the power as a function of β while varying T , whereas in Figures 2 and 3, we keep T fixed and instead consider varying r and r_1 . For simplicity, $N = 20$ is kept fixed and ρ is set to zero. The bandwidth is set according to the Newey and West (1994) data dependent rule. As expected, we see that the power is increasing in T and in the distance from the null, as measured by $|\beta|$. The best power is obtained when $r = r_1 = 1$, which is to be expected because as long as $r_1 \geq 1$ the test does not make use of the fact that there may be more than one unit root. In addition, we see that the rate of consistency of $\hat{\omega}^2$ is not only affecting the size of the test but also the power, especially when β is close to its hypothesized value under the null. In particular, we see that the power of

requirement. The relatively high rate of consistency in N is due to the fact that the estimation is based on the cross-sectional averages \bar{X}_t rather than X_{it} itself.

¹³Unreported simulation results show that the test tends to be severely undersized when $r_1 = 0$, which confirms our theoretical results.

Figure 2: Power when $T = 50$.

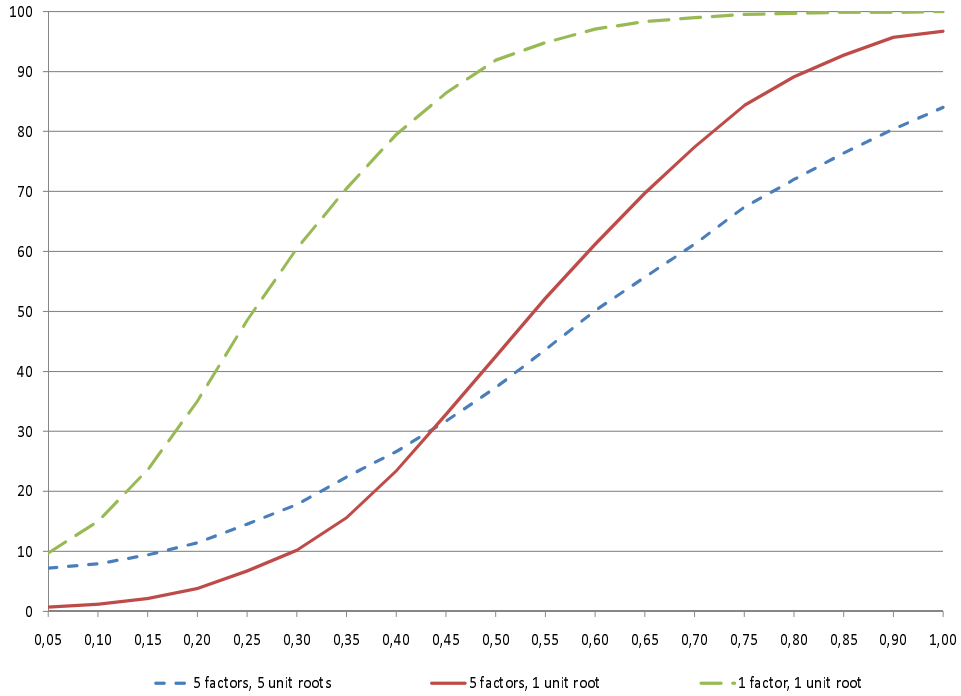
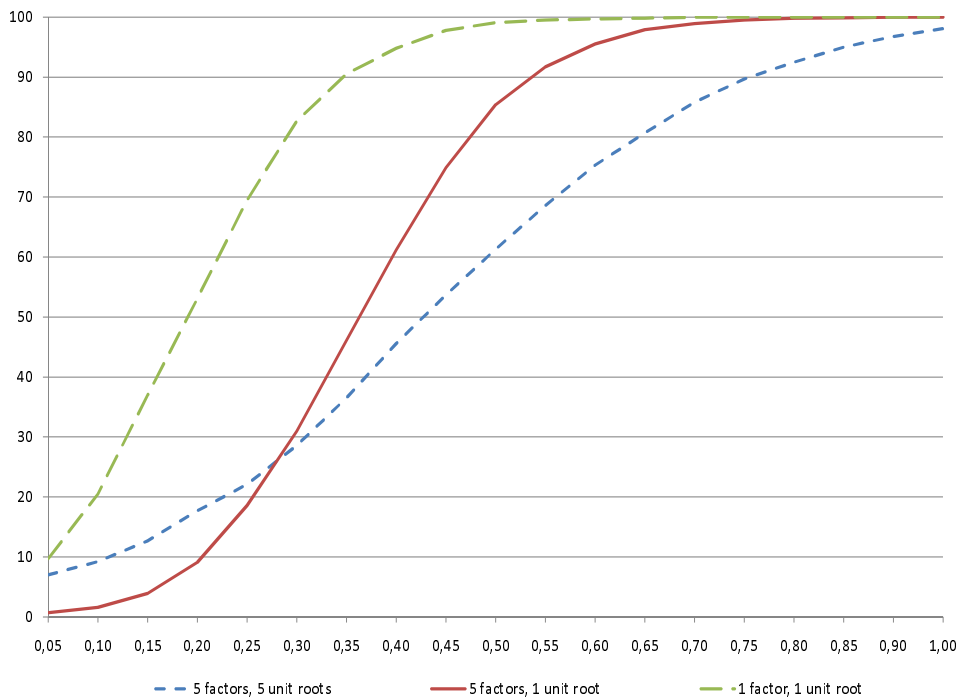


Figure 3: Power when $T = 100$.



the test based on five factors drops substantially as the number of unit roots goes from five to one. As T increases, however, the difference gradually disappears.

In summarizing this section we find that the new test has a number of distinct features that makes it very attractive from both an applied and a theoretical point of view. Firstly, the test can be applied with little prior knowledge regarding the degree of integration of the common and idiosyncratic components. The only restriction is that there must be at least one unit root factor present, which is of course a testable restriction. Secondly, the test is robust against quite general forms of serial and cross-sectional dependence, and still it requires only minimal corrections. In fact, as for the cross-sectional dependence, as long as it has a common factor structure with at least one unit root, then there is no need for any correction at all. Thirdly, the test has good finite sample properties with small size distortions and high power even when N is as small as 20 and T is as small as 50.¹⁴

3.3 A sequential PANIC procedure

The above discussion suggests that if the data contain a constant, as is usually the case, but there is uncertainty about the trend, then the following sequential procedure can be used.¹⁵

1. Obtain \hat{F}_t and \hat{e}_{it} by applying the principal components method to $\Delta X_{it} - \overline{\Delta X}_i$.
2. Test for unit roots in \hat{F}_t using the ADF test.¹⁶
3. If the null of a unit root is rejected for all the elements of \hat{F}_t at Step 2, we conclude that F_t is stationary and continue to test for unit roots in \hat{e}_{it} using the P test.
 - a. If the null of a unit root is rejected, we conclude that e_{it} , and therefore also X_{it} , is stationary, and proceed no further. The significance of the trend can now be tested by using standard techniques for stationary data.
 - b. However, if the null is accepted, then we conclude that e_{it} has at least one unit root and therefore so must X_{it} , and so the procedure is stopped.

¹⁴We also ran some simulations with $N = 21$ and $T = 31$, which is the sample size considered here, but with no major changes to the results.

¹⁵See Ayat and Burrige (2000) for a similar procedure in the pure time series context.

¹⁶One way to control the overall significance level at Step 2 in case of multiple factors is to use a multivariate rank test, such as the trace test of Johansen (1995) or the MQ tests of Bai and Ng (2004). However, unreported simulation results suggest that the choice of test at this step has little or no effect on the performance of the sequential PANIC procedure. In fact, in samples of the same size as the one considered here, the ADF test actually performed best.

4. If the null is accepted for at least one of the elements of \hat{F}_t at Step 2, then we proceed to test for the significance of the trend using the t_β test.
5. If the null of no trend is rejected at Step 4, then \hat{e}_{it} is tested for unit roots, again using the P test.
 - a. If the unit root null is rejected, we conclude that the non-stationarity of X_{it} is due to the common component, and stop the procedure.
 - b. On the other hand, if the null is accepted, then we conclude that the non-stationarity of X_{it} is due to both components, and stop the procedure.
6. If the null of no trend is accepted at Step 4, F_t and e_{it} reestimated by applying the principal components method to ΔX_{it} .
7. The estimated components from Step 6 are tested for unit roots using the standard PANIC approach in the absence of a trend.

As pointed out earlier the main dilemma here is that while we would like to be able to increase the power of the unit root tests by removing the trend, by doing so we run the risk of obtaining biased results that will make it difficult to reject the unit root null even when it is false. The above procedure is designed to minimize the risk of such bias.

As with all other sequential unit root tests with a possible trend, the behavior of the above PANIC procedure reflects the interplay between test power and size at each step (Ayat and BurrIDGE, 2000). Consider for example a situation in which the ADF test in Step 2 is correctly sized but has low power. If the null of a unit root is rejected, then we are confident that there is no lack of power, and so the inclusion of the potentially superfluous trend is not a problem. There is therefore no need to apply the t_β test, which we know is invalid when all the factors are stationary. On the other hand, if the unit root null is not rejected, then we do not know if this is due to genuine non-stationarity, or simply the poor power of the test. Since the latter is going to make the t_β test in Step 4 biased towards the null of no trend, the probability of correctly identifying trend-stationary common factors is reduced. Similarly, if the ADF test in Step 2 is oversized, then the likelihood is that the procedure is going to label the factors as stationary when in fact they are non-stationary.

Fortunately, the simulation evidence suggests that our procedure is not seriously affected by these problems. This is shown in Table 3, where we report the 5% size and power of the

sequential factor and idiosyncratic unit root tests. For simplicity we use the same simulation design as in Table 1. In particular, because the results for the case with $r \geq 2$ were basically the same, we assume that there is a single factor, and to be able to isolate the effect of the trend test, we further assume that there is no serial correlation. As in Table 2, the bandwidth is set using the Newey and West (1994) rule.

Two values for the assumed homogenous trend slope β are considered, zero and one. However, the procedure was almost always able to detect the trend when present in the non-stationary factor case, and in the stationary factor case the ADF test rejected, causing the procedure to stop in the trend model. The results for the case when $\beta = 1$ were therefore almost identical to those reported in Table 1 based on taking the trend as given. For this reason, in Table 3 we only report the results for the case when $\beta = 0$. Because the constant-only tests are now correct while the trend tests include a superfluous deterministic regressor, the former is expected to perform best, with the sequential tests lying somewhere in between. In fact, this is exactly what we see in the table. Note in particular how the performance of the constant-only tests is only slightly better than that of the sequential tests, suggesting that the t_β test is performing well.

4 Empirical Results

4.1 Data

The data we use are annual and cover the 21 Swedish counties between 1975 and 2008. The crime rates are defined as the number of reported offences to the police per 100,000 of the population.¹⁷ Two crime categories are considered, property and violent crimes. We will focus on two of the most common property crimes, burglary and theft. Regarding violent crimes, most of the previous studies have considered robbery and homicide, and therefore so do we.¹⁸ A more detailed description of the data and our sources is given in Appendix B.

¹⁷While we would like to use data on crimes actually committed, there are good reasons for why the number of reported offences to the police is a good measure of this. For example, consider property crimes. Since reporting the crime is necessary for receiving insurance compensation, the error incurred when replacing actual offences by reported offences is likely to be small.

¹⁸Although there is no consensus about this, in the present study we regard robbery as a violent crime.

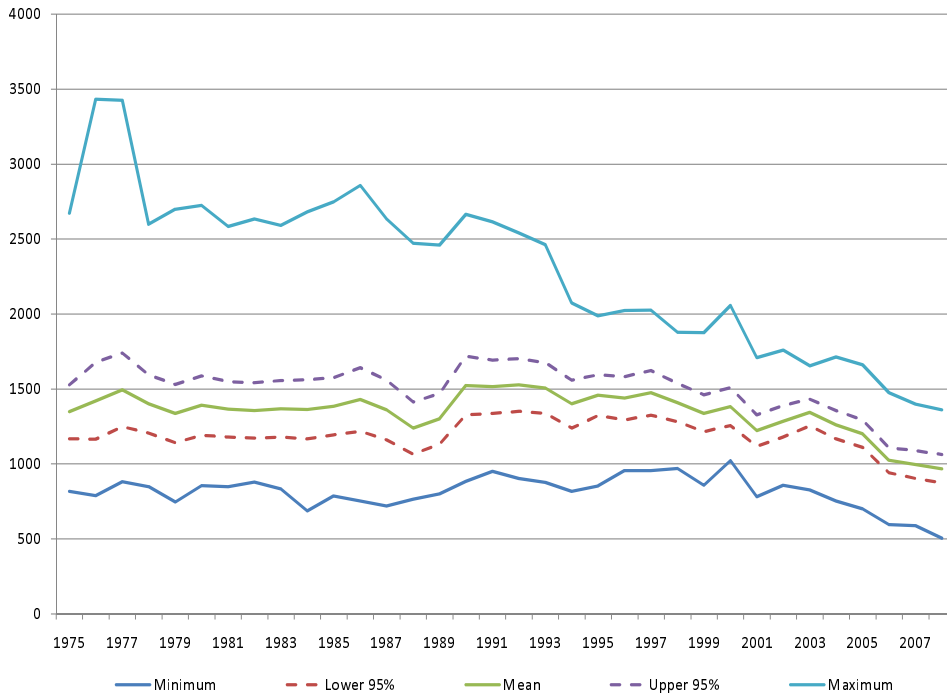
4.2 Preliminary evidence

In order to get a feeling for the persistence and cross-correlation of the different crimes, we begin with a graphical inspection of the data. Figures 4 through 7 plot the cross-regional mean, range and normal 95% confidence bands for each of the four crime types. In Section 3.2 we illustrated how $\overline{\Delta X}_t$ can be used to approximate ΔF_t . A similar result holds for \overline{X}_t . In particular,

$$\overline{X}_t = \overline{D}_t + \overline{\lambda}' F_t + \overline{e}_t = \overline{D}_t + \overline{\lambda}' F_t + o_p(1),$$

suggesting that \overline{X}_t can be regarded as a measure of the common component of crime, which should not have unit roots if the regional crime rates are stationary. However, the figures show no evidence of mean reversion, suggesting that the common components of all four crimes are non-stationary. Hence, we cannot rule out the possibility that crime rates may be cointegrated across counties.

Figure 4: Cross-regional mean, range and confidence bands of burglary.



We also see that the mean is able to explain a large part of the overall variation in the data. To take one example, if we look at theft in Figure 5 there is an upward trend during the

Figure 5: Cross-regional mean, range and confidence bands of theft.

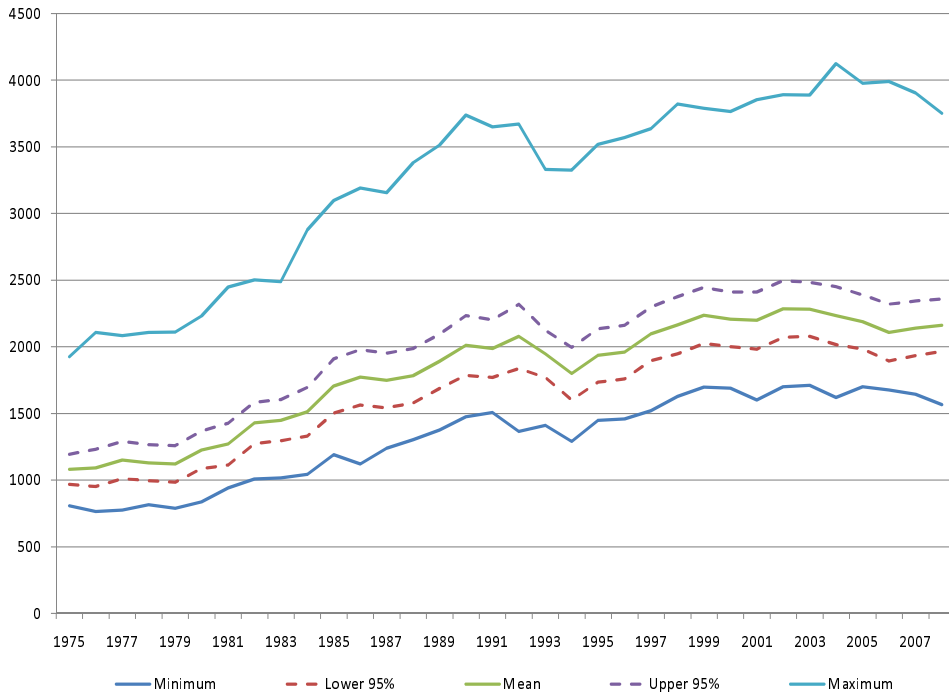


Figure 6: Cross-regional mean, range and confidence bands of robbery.

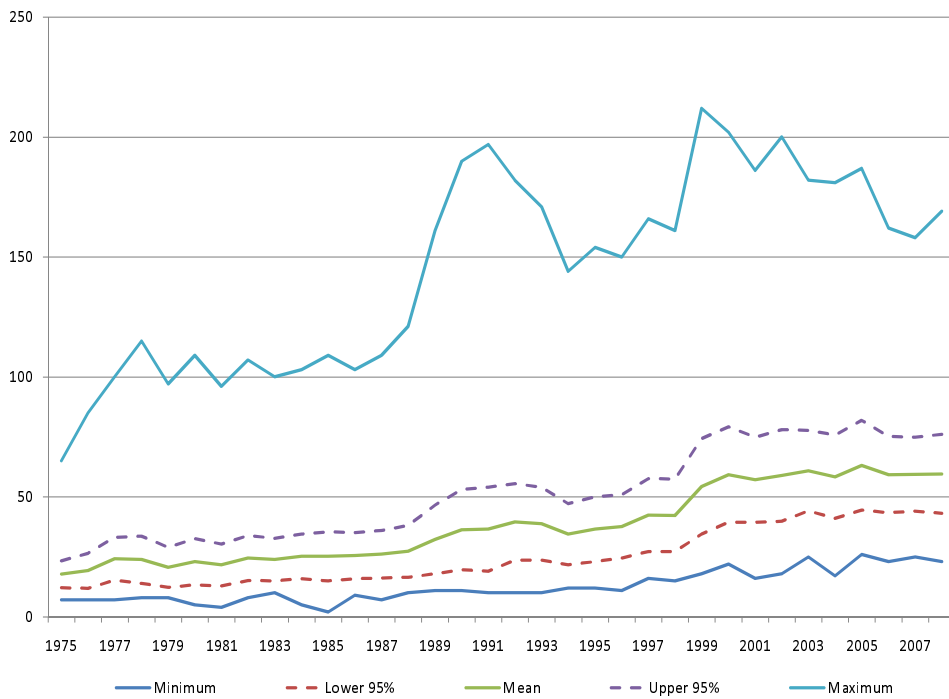
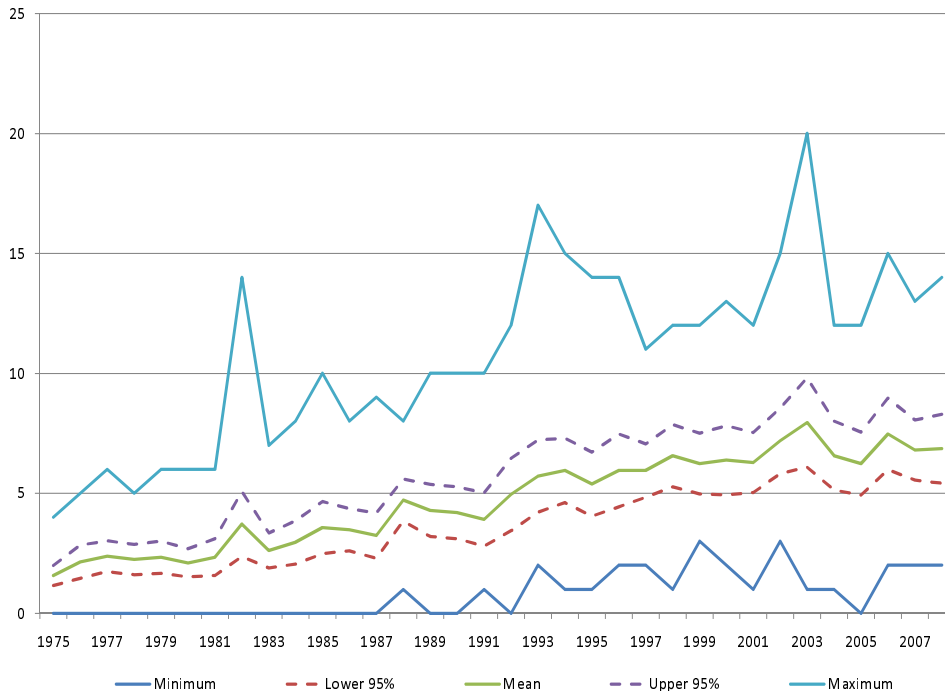


Figure 7: Cross-regional mean, range and confidence bands of homicide.



whole period except in the early 1990's when theft declined. However, while trending, the series do not drift far apart. Thus, the common component to theft seem to be rather strong.

Of course, although useful for developing a feeling for the degree of mean reversion, graphical evidence of this sort does not provide any statistical evidence of whether the county-level crime rates are actually stationary or not. This is where the PANIC method comes in, the results of which are reported in section 4.3.

In order to infer the statistical significance of the cross-correlations, we compute the pairwise cross-county correlation coefficients of each of the first differenced crime variables. The simple average of these correlation coefficients across all the 210 county pairs, together with the associated CD test discussed in Pesaran *et al.* (2008), are given in Table 4. The average correlation coefficients are very high, between 0.85 and 0.99, and the CD statistics are highly significant, which obviously strengthens the case against independence. Thus, as expected, crime rates across counties are not independent of each other. One implication of this is that the first generation of panel unit root tests used by for example in Moody and Marvell (2005) and Phillips (2006) are likely to be deceptive, and that the use of PANIC is more appropriate.

4.3 PANIC

The preliminary results reported so far indicate that at least some of the crime rates may be non-stationary. To investigate the statistical significance of these results, we now proceed to discuss the results from the sequential PANIC procedure of Section 3.3. We begin by looking at the results from the estimation and testing of F_t , which are then used in determining the significance of the trend. Finally, we take a look at the results for the estimated idiosyncratic component.

Following the recommendation of Bai and Ng (2002), the number of factors is determined using the IC_{p2} information criterion. The maximum number of factors is set to six.¹⁹ For robbery and theft we end up with five and two factors, respectively, while for burglary and homicide we estimate one factor.

Table 5 reports the ADF test results for each of the factors, where the lag length has been determined using the Schwarz Bayesian information criterion. The first thing to notice is that the results differ depending on whether there is a constant or a constant and trend in the model. Thus, just as discussed in Section 3.3 the decision of whether to include the trend or not is going to play an important role here. Of course, at this stage we do not know if the trend can be safely removed and so we look at the results with the trend included. The 5% critical value for the ADF test is -3.41 , which leads to at least one acceptance for each crime, suggesting that the common components of all four crimes are non-stationary.

We also see that the estimated factors account for a large fraction of the variance in the panel, with the first factor accounting for between 20% and 35% of the total variation.²⁰ Together the five factors of robbery account for more than 75% of the total variation, which represents the largest common component. Homicide have the smallest common component with only one factor that accounts for about 20% of the total variation.

The results obtained from applying the trend test are reported in Table 6. We see that the slope coefficients for theft, robbery and homicide are significant at the 5% level suggesting that for these crimes we should keep the trend in the model. Thus, looking again at Table 5, and the trend results reported therein, we see that among the five factors of robbery there

¹⁹Since our panel is quite small, we do not consider more than six factors, as this will only lead to imprecise factor estimates.

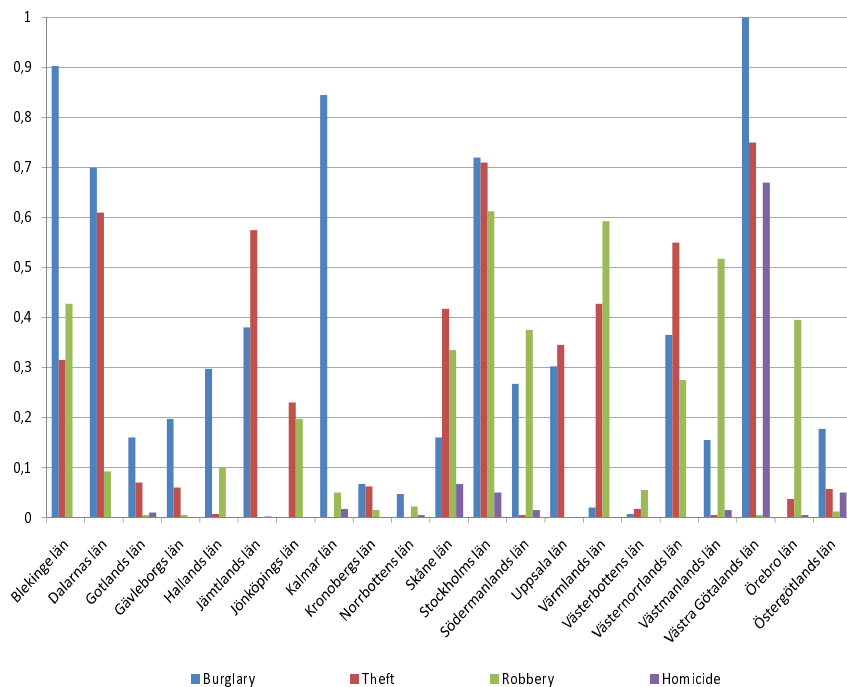
²⁰The first factor explains the largest fraction of the total variation in the panel, while the second factor explains the largest fraction of the variation controlling for the first factor, and so on. The estimated factors are mutually orthogonal by construction.

are four instances where the null of a unit root cannot be rejected at the 5% level. Regarding theft and homicide, all factors are non-stationary. In other words, for these crimes there is evidence not only of deterministic trends but also of common stochastic trends.

For burglary, however, the trend is insignificant and can therefore be removed. The ADF test results in Table 5 for the case with a constant but no trend shows that the null of a unit root cannot be rejected at the 5% level, which is in agreement with the result for the trend case. It follows that the common components of all four crimes are non-stationary. Moreover, while theft, robbery and homicide are trending deterministically, burglary is not.

With this in mind we now proceed to test for unit roots among the estimated idiosyncratic components. The results from the Bai and Ng (2004) $P_{\hat{\epsilon}}$ test are reported in Table 7, where we have again made use of the Schwarz Bayesian criterion for determining the order of the lag augmentation. It is seen that the evidence is uniformly against the unit root null, and we therefore conclude that the idiosyncratic component of each crime category is stationary for the panel as a whole.

Figure 8: County-specific unit root test p -values.



Of course, the fact that the panel test rejects does not mean that the crime rate of each

individual county is stationary.²¹ This is seen in Figure 8, which plots the p -values obtained by applying the ADF test to each county. Looking at the 10% level, we see that the null is rejected 20 times for homicide, 12 times for robbery, 10 times for theft and six times for burglary. In other words, while still rather strong, as expected the evidence of stationarity at the individual county-level is weaker than at the aggregate panel level. In any case, since we cannot reject the presence of a unit root in the common components, all four crimes must be considered as non-stationary. The presence of non-stationary factors and stationary idiosyncratic components means that crime rates are cointegrated across counties.

As mentioned in Section 3.1 cross-county cointegration is similar to the concept of club convergence, see for example Bernard and Durlauf (1995, 1996) for a formal definition. But actually cross-county cointegration is just a necessary condition for club convergence. To appreciate this, suppose that F_t is non-stationary and that e_{it} is stationary. According to the definition of Bernard and Durlauf (1995, 1996), X_{it} and X_{jt} form a convergence club if their difference is stationary. But

$$X_{it} - X_{jt} = D_{it} - D_{jt} + (\lambda_i - \lambda_j)'F_t + e_{it} - e_{jt},$$

suggesting that for $X_{it} - X_{jt}$ to be stationary we also require $\lambda_i = \lambda_j$ for all i and j . Cross-county cointegration alone is therefore not enough for convergence.

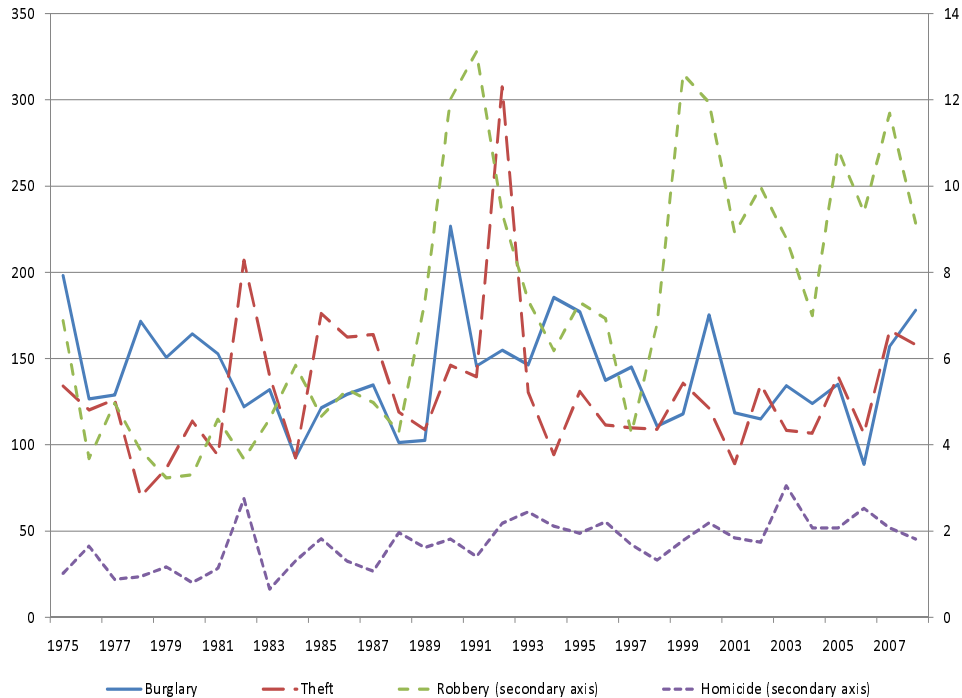
Of course, this paper is not about convergence, and it is not our intention here to sort out how and why some counties are convergent while others are not. Nevertheless, to get a feeling for how reasonable the convergence restriction of equal loadings is, we estimated by least squares a pooled regression of $\Delta X_{it} - \Delta X_{jt}$ on a constant and $\Delta \hat{F}_t$. If the loadings are equal R^2 should be close to zero and the factors should be insignificant.²² In agreement with this, we find that the R^2 statistics from the four regressions lie between zero and 0.04, and the F -test for the exclusion restriction of all the factors results in two non-rejections at the 1% significance level.

Similar conclusions can be reached by looking at Figure 9, which plots the cross-county standard deviation for each crime after detrending, which is a common measure of convergence. As pointed out by Pesaran (2007b), if the loadings are equal, then this measure

²¹Strictly speaking, for P_e to end up in a rejection of the null of a unit root in the idiosyncratic component of all 21 counties it is enough that the idiosyncratic component of one of the counties is stationary.

²²See Section 4.4 for a theoretical motivation of this regression, which uses the estimated factors as generated regressors.

Figure 9: Crime cross-county standard deviations after detrending.



should not have a unit root or be trending. The obvious mean-reversion of all four series is supported by the ADF test.²³

4.4 The importance and interpretation of the factors

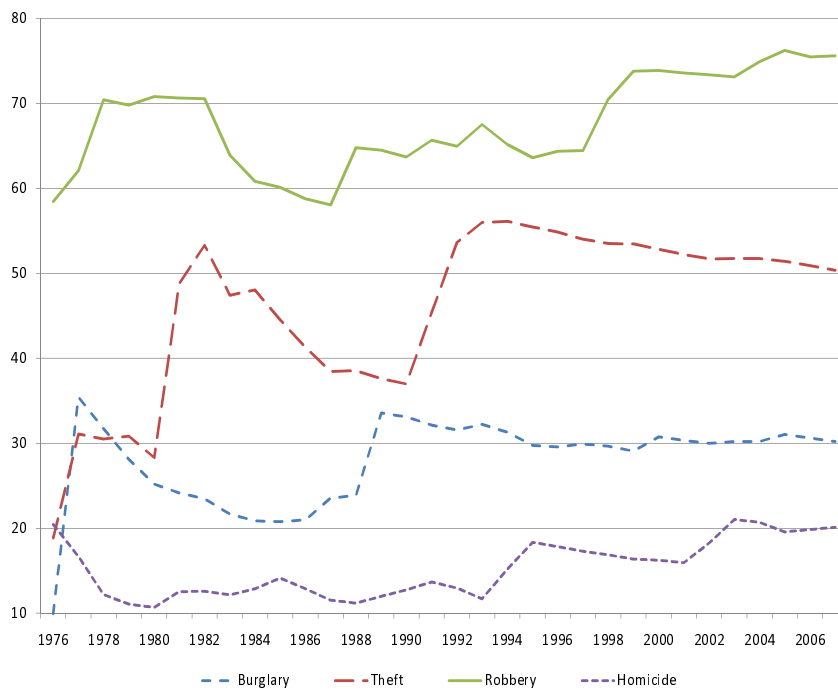
As an illustration of how the importance of the factors has changed over time, Figure 10 plots the fraction of the total variation in the data that can be explained by the estimated common component.²⁴ The first thing to notice is the similarity with which the common components have developed over time. The importance of the common shocks changed dramatically during the first half of the sample, a period largely consistent with the turbulence of the late 1980's, and the overheating of the Swedish economy. The importance of the common shocks then starts to stabilize, levelling off in the end of the sample, which is also something that is reflected in the macroeconomic data. The deep recession that followed the overheating

²³The results look different when the data are not detrended, which is not totally unexpected given that we have already established that theft, robbery and homicide contain trends. Thus, although X_{it} and X_{jt} seem to be cointegrated, they may not be cotrending. On the other hand, the estimated intercepts in our pooled regressions are insignificant, suggesting that both types of trends cancel out in the pair-wise crime differences.

²⁴To guard against spurious effects, the variance is calculated from the first-differenced data.

of the late 1980's persisted for quite a while but then it started to fade out. In terms of real output the recovery was quick, but the unemployment rate remained high until the late 1990's. In agreement with the results of Table 4 we also see that the importance of the common component is largest with robbery, and that it is smallest with homicide.

Figure 10: The fraction of the total variance explained by the estimated common component.



Given the importance common components, it is interesting to consider the driving forces behind the estimated factors. The results of the previous literature suggest that crime is driven by factors such as unemployment and income. In this section we therefore make an attempt to label the estimated factors according to their relationship with macroeconomic variables.

This is done by regressing each of the factors onto a small set of macroeconomic country-level variables, including unemployment, per-capita private consumption and per-capita gross domestic product (GDP).²⁵ The ADF test indicate that the unit root null is accepted at the 5% level for all three variables. Therefore, since both the factors and explanatory variables seem to be contaminated with unit roots, in order to minimize the risk of obtaining spurious regression results, we work with first differences rather than levels. Lagged values

²⁵See Appendix B for a more detailed description of the data.

of both the first-differenced factors and regressors were included if it improved the fit if the regression, as measured by the Schwarz Bayesian criterion.

It should be noted that the dependent variable here is $\Delta\hat{F}_t$, which is an estimate of ΔF_t . Thus, since we are dealing with a generated dependent variable, one might inquire as to the validity of the resulting regression. The following argument can be used. Write

$$\Delta F_t = AZ_t + u_t,$$

where Z_t is the vector of contemporaneous and possibly also lagged explanatory variables, A is a matrix of slope coefficients and u_t is a mean zero stationary error term. Hence, by pre-multiplication of the rotation matrix H , $H\Delta F_t = HAZ_t + Hu_t$, and then adding and subtracting $\Delta\hat{F}_t$,

$$\Delta\hat{F}_t = HAZ_t + Hu_t + (\Delta\hat{F}_t - H\Delta F_t) = BZ_t + w_t.$$

It follows that

$$\hat{B} - B = \left(\sum_{t=2}^T Z_t Z_t' \right)^{-1} \sum_{t=2}^T Z_t (Hu_t + (\Delta\hat{F}_t - H\Delta F_t)),$$

where $\frac{1}{\sqrt{T}} \sum_{t=2}^T Z_t (\Delta\hat{F}_t - H\Delta F_t) = o_p(1)$ when $\frac{\sqrt{T}}{N} \rightarrow 0$, see Lemma A.1 of Bai and Ng (2006). Hence,

$$\sqrt{T}(\hat{B} - B) = \left(\frac{1}{T} \sum_{t=2}^T Z_t Z_t' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=2}^T Z_t Hu_t + o_p(1),$$

which shows that the effect of the estimated dependent variable is negligible. However, because A is not identified after replacing ΔF_t with $\Delta\hat{F}_t$, the sign of the estimated coefficients have no particular meaning. In Table 8 we therefore only report the p -values for each variable. More precisely, the p -values are for the exclusion restriction of both the contemporaneous and lagged values of each of the explanatory variables. If the model includes lagged values of the dependent variable, then the p -value for the exclusion restriction of these lags is also reported. The standard errors are estimated using the Newey and West (1994) procedure. For each regression two measures of the overall fit are reported, the R^2 statistic and the p -value of an F -test of the null hypothesis that all the explanatory variables but the constant can be excluded.

Starting with the violent crimes, we see that the common factor of homicide, having the highest R^2 of 73%, loads from both unemployment and private consumption. Robbery is

also related to the macroeconomy. Specifically, while factors three and five load mainly from unemployment, the second factor loads mainly from unemployment and per-capita GDP.

Turning next to property crimes, which have received most attention in the literature, we see that the common factor of burglary is largely unexplained. The R^2 statistic is low and we cannot reject the null that the coefficients of all three regressors are jointly zero. The results for theft are more promising with the second factor loading significantly from private consumption. As for the first factor we find that although the p -values of the regressors are individually insignificant, the F -test clearly rejects that they are unimportant, which is typical sign of multicollinearity. The R^2 statistic is almost as high as for homicide, around 71%. Moreover, since this factor accounts for about 30% of the total variation in the data, it is clear that the economy is an important determinant of theft.

In summary, for three out of the four crimes considered we find a significant association between the common factors and the macroeconomic conditions. However, even if these factors seem to be interpretable, we would like to point out that the results do not say anything about the strength and direction of the association.

4.5 Robustness

As we have argued above, the PANIC approach used here is very robust in the sense that it permits not only for county specific deterministic terms and serial correlation but also for a wide range of cross-regional interdependencies, including dependence in the form of cross-county cointegration.

One weakness is that the above analysis does not allow crime to be structurally shifting. Although Figures 4 through 7 do not lend much support of such shifts, we would still like to allow for the possibility that there might be. In order to investigate this issue more formally, we employ a version of the procedure of Perron and Rodríguez (2003), which is based on testing for breaks in the first-differenced data. Suppose for example that there is a break in the level of X_{it} such that $D_{it} = c_i + \beta_i t + \delta_i B_{it}$, where B_{it} is a dummy variable that represents the location of the structural break for each county. It follows that

$$\Delta X_{it} = \beta_i + \delta_i \Delta B_{it} + \lambda_i' \Delta F_t + \Delta e_{it}.$$

As pointed out by Perron and Rodríguez (2003), writing the model in differences in this way is advantageous for at least three reasons. Firstly, since ΔX_{it} is stationary, the procedure

is robust against possible unit roots in F_t and e_{it} . Secondly, since the level break is now an outlier, the breakpoint can be estimated by using conventional methods for outlier detection. Thirdly, there may be multiple breaks.²⁶

Applying this procedure to our data, we find only two violations of the no break null, one for theft and one for robberies. Thus, there seem to be very little evidence of structural instability. Moreover, redoing the analysis while conditioning on the estimated breaks, we reach exactly the same conclusions as before.

5 Concluding remarks

In this study, we try to shed some light on the persistence and interregional dependency of crime, an often neglected feature of empirical studies of the economics of crime. For this purpose, the PANIC methodology of Bai and Ng (2004) is employed, which enables us to first estimate and then to test for unit roots in two components of the data, an idiosyncratic component and a common component. This decomposition is appropriate because crime rates usually exhibit both high variability within each region over time and strong comovements across regions, features that are consistent with the recent theoretical models of crime. The problem is that PANIC assumes that the researcher knows whether a deterministic trend is present or not, which is not very realistic. The current paper therefore develops a sequential PANIC procedure that determines the extent of both the trend and the non-stationarity of the data.

Using a panel that covers 21 Swedish counties between 1975 and 2008, we are able to reject the presence of a unit root in the estimated idiosyncratic component for all four crimes considered but not in the estimated common component. Specifically, we find that all common components have at least one unit root, which leads us to the conclusion that the crimes are cointegrated across counties. The fact that these components are also relatively important suggests that most crime shocks are common. Thus, according to our results crime shocks are not likely to dissipate with time but are more likely to be persistent and to spread across counties, just as predicted by theory.

One implication of this result is that the conventional approach of employing conventional regression techniques designed for stationary panels may be hazardous. It also sug-

²⁶The main disadvantage with this procedure is that it does not allow for the possibility of breaks in the trend. However, in our data this is not too much of a problem, because there are no apparent trend breaks.

gests that the conclusions from prior research need to be reevaluated, as the possibility remains that they have been spuriously induced by the presence of cross-unit cointegration. This is a potentially very serious issue, as nearly all of the leading studies in the field assume that the data are stationary.

Another implication is that since most previous studies do not account for both the dynamics and the cross-correlations of the data, they are likely to misstate the effects of current shocks on future crime rates.

As far as policy is concerned, because the non-stationarity seem to originate with nationwide factors rather than with county-specific disparities, our results indicate that crime preventing actions should be centralized. In particular, policy actions aimed at lowering violent crimes should focus on dampening the effects of unemployment.

References

- Ayat, L., Burridge, P. (2000). Unit root tests in the presence of uncertainty about the non-stochastic trend. *Journal of Econometrics* **95**, 71–96.
- Bai, J., Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191–221.
- Bai, J., Ng, S. (2004). A PANIC attack on unit roots and cointegration. *Econometrica* **72**, 1127–1177.
- Bai, J., Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* **74**, 1133–1150.
- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy* **76**, 169–217.
- Bernard, A. B., Durlauf, S. (1995). Convergence in international output. *Journal of Applied Econometrics* **10**, 97–108.
- Bernard, A. B., Durlauf, S. (1996). Interpreting tests of the convergence hypothesis. *Journal of Econometrics* **71**, 161–173.
- Breitung, J., Pesaran, M. H. (2008). Unit roots and cointegration in panels. In Matyas, L., and Sevestre, P. (Eds.), *The econometrics of panel data: Fundamentals and recent developments in theory and practice*, 279–322. Kluwer Academic Publishers: Boston.
- Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance* **20**, 249–272.
- Dickey, D. A., Fuller, W. A. (1979). Distribution of the estimator for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427–431.
- Edmark, K. (2005). Unemployment and crime: Is there a connection? *Scandinavian Journal of Economics* **107**, 353–373.
- Fajnzylber, P., Lederman, D., Loayza, N. (2002). What causes violent crime? *European Economic Review* **46**, 1323–1357.

- Forni, M., Lippi, M. (1997). *Aggregation and the microfoundations of dynamic macroeconomics*. Oxford University Press: Oxford.
- Funk, P., Kugler, P. (2003). Dynamic interactions between crimes. *Economics Letters* **79**, 291–298.
- Glaeser, E. L., Sacerdote, B., Scheinkman, J. A. (1996). Crime and social interactions. *Quarterly Journal of Economics* **111**, 507–548.
- Gould, E. D., Weinberg, B. A., Mustard, D. B. (2002). Crime rates and local labor market opportunities in the United States: 1979–1997. *Review of Economics and Statistics* **84**, 45–61.
- Gutierrez, L. (2006). Panel unit-root tests for cross-sectionally correlated panels: A Monte Carlo comparison. *Oxford Bulletin of Economics and Statistics* **68**, 519–540.
- Hale, C. (1998). Crime and the business cycle in the post-war Britain revisited. *British Journal of Criminology* **38**, 681–698.
- Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.
- Levitt, S. D. (1999). The limited role of changing age structure in explaining aggregate crime rates. *Criminology* **37**, 581–598.
- McDowall, D., Loftin, C. (2009). Do US city crime rates follow a national trend? The influence of nationwide conditions on local crime patterns. *Journal of Quantitative Criminology* **25**, 307–324.
- Moody, C. E., Marvell, T. B. (2005). Guns and crime. *Southern Economic Journal* **71**, 720–736.
- Murphy, K. M., Shleifer, A., Vishny, R. W. (1993). Why is rent-seeking so costly to growth? *American Economic Review* **83**, 409–414.
- Newey, W. K., West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* **61**, 613–653.
- Perron, P., Rodríguez, G. (2003). Searching for additive outliers in nonstationary time series. *Journal of Time Series Analysis* **24**, 193–220.

- Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74**, 967–1012.
- Pesaran, M. H. (2007a). A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics* **22**, 265–312.
- Pesaran, M. H. (2007b). A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics* **138**, 312–355.
- Pesaran, H. M., Ullah, A., Yamagata, Y. (2008). A bias-adjusted LM test of error cross section independence. *Econometrics Journal* **11**, 105–127.
- Phillips, J. A. (2006). The relationship between age structure and homicide rates in the United States, 1970–1999. *Journal of Research in Crime and Delinquency* **43**, 230–260.
- Raphael, S., Winter-Ebmer, R. (2001). Identifying the effect of unemployment on crime. *Journal of Law and Economics* **44**, 259-283.
- Sah, R. (1991). Social osmosis and patterns of crime. *Journal of Political Economy* **99**, 1272–1295.
- Westerlund, J. (2009). A note on the use of the LLC panel unit root test. Forthcoming in *Empirical Economics*.
- Witt, R., Clarke, A., Fielding, N. (1998). Common trends and common cycles in regional crime. *Applied Economics* **30**, 1407–1412.

Appendix A: The asymptotic distribution of the trend test

In this appendix we prove (9). We begin by assuming that ΔF_t and Δe_{it} are serially uncorrelated, and that all the elements of F_t and e_{it} are non-stationary. We then show how the results generalize to cases when these assumptions do not hold.

Note first that since e_{it} is non-stationary,

$$\sqrt{NT} \bar{\Delta e} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=2}^T \epsilon_{it},$$

which is $O_p(1)$ if e_{it} is also independent across i . To show this, since ϵ_{it} is mean zero, we only need to verify that the variance is finite so that a central limit theorem applies. Letting $\tau_{ij} = E(\epsilon_{it}\epsilon_{jt})$ denote the cross-sectional covariance of ϵ_{it} , we obtain

$$\text{var}(\sqrt{NT} \bar{\Delta e}) = \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=2}^T \sum_{s=2}^T E(\epsilon_{is}\epsilon_{jt}) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \tau_{ij},$$

which is zero under cross-sectional independence. Under weak cross-section dependence, $\sum_{i=1}^N \tau_{ij}$ is not necessarily zero but finite, see Assumption C of Bai and Ng (2004). Suppose therefore that $\sum_{i=1}^N \tau_{ij} \leq K$, where K is finite, in which case we obtain

$$\text{var}(\sqrt{NT} \bar{\Delta e}) = \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \tau_{ij} \leq \frac{1}{N} \sum_{j=1}^N K = K,$$

showing that the order of $\sqrt{NT} \bar{\Delta e}$ is not affected by allowing for weak cross-section dependence in ϵ_{it} . By using this result together with the definition of $\hat{\beta}$ and (8),

$$\sqrt{T}(\hat{\beta} - \bar{\beta}) = \bar{\lambda}' \sqrt{T} \bar{\Delta F} + \sqrt{T} \bar{\Delta e} = \bar{\lambda}' \sqrt{T} \bar{\Delta F} + O_p\left(\frac{1}{\sqrt{N}}\right),$$

where $E(\bar{\lambda}' \sqrt{T} \bar{\Delta F}) = \bar{\lambda}' \sqrt{T} E(\bar{\Delta F}) = 0$ and

$$\text{var}(\bar{\lambda}' \sqrt{T} \bar{\Delta F}) = \bar{\lambda}' \text{cov}(\sqrt{T} \bar{\Delta F}) \bar{\lambda} = \bar{\lambda}' \frac{T}{(T-1)^2} \sum_{t=2}^T \text{cov}(\Delta F_t) \bar{\lambda} \rightarrow_p \lambda' \Sigma \lambda$$

as $N, T \rightarrow \infty$, where $\lambda = \lim_{N \rightarrow \infty} \bar{\lambda}$ and $\Sigma = \text{cov}(\Delta F_t)$. Hence, by applying the Lindeberg–Levy central limit theorem to the first term,

$$\sqrt{T}(\hat{\beta} - \bar{\beta}) = \bar{\lambda}' \sqrt{T} \bar{\Delta F} + O_p\left(\frac{1}{\sqrt{N}}\right) \rightarrow_d \sqrt{\lambda' \Sigma \lambda} N(0, 1).$$

The standard error of $\hat{\beta}$ is given by $\frac{\hat{\sigma}}{\sqrt{T}}$, where $\hat{\sigma}^2$ can be expanded as follows:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{T-1} \sum_{t=2}^T (\overline{\Delta X}_t - \hat{\beta})^2 = \frac{1}{T-1} \sum_{t=2}^T (\bar{\lambda}' (\Delta F_t - \overline{\Delta F}) + \overline{\Delta e}_t - \overline{\Delta e})^2 \\
&= \bar{\lambda}' \frac{1}{T-1} \sum_{t=2}^T (\Delta F_t - \overline{\Delta F}) (\Delta F_t - \overline{\Delta F})' \bar{\lambda} \\
&\quad + \bar{\lambda}' \frac{2}{\sqrt{N(T-1)}} \frac{1}{\sqrt{T-1}} \sum_{t=2}^T (\Delta F_t - \overline{\Delta F}) \sqrt{N} (\overline{\Delta e}_t - \overline{\Delta e}) \\
&\quad + \frac{1}{N} \frac{1}{T-1} \sum_{t=2}^T N (\Delta e_t - \overline{\Delta e})^2 \\
&= \bar{\lambda}' \frac{1}{T-1} \sum_{t=2}^T (\Delta F_t - \overline{\Delta F}) (\Delta F_t - \overline{\Delta F})' \bar{\lambda} + O_p \left(\frac{1}{\sqrt{NT}} \right) + O_p \left(\frac{1}{N} \right) \\
&= \bar{\lambda}' \hat{\Sigma} \bar{\lambda} + O_p \left(\frac{1}{\sqrt{NT}} \right) + O_p \left(\frac{1}{N} \right),
\end{aligned}$$

from which follows that $\hat{\sigma}^2 \rightarrow_p \lambda' \Sigma \lambda$ as $N, T \rightarrow \infty$, and so we obtain

$$t_\beta = \frac{\sqrt{T}(\hat{\beta} - \bar{\beta})}{\hat{\sigma}} \rightarrow_d N(0, 1).$$

The above result holds even if Δe_{it} is serially correlated. However, if the correlation is in ΔF_t , things change. In particular,

$$\text{cov}(\bar{\lambda}' \sqrt{T} \overline{\Delta F}) \rightarrow_p \lambda' \lim_{T \rightarrow \infty} \text{cov}(\sqrt{T} \overline{\Delta F}) \lambda = \lambda' \Omega \lambda,$$

where $\Omega = \sum_{s=-\infty}^{\infty} E(\Delta F_0 \Delta F_s')$ is the long-run covariance of ΔF_t , suggesting that for t_β to be asymptotically standard normal $\hat{\sigma}^2$ in the denominator needs to be replaced by a consistent estimator of the long-run variance of $\bar{\lambda}' \sqrt{T} \overline{\Delta F}$.

Let us decompose Ω as

$$\Omega = \Sigma + \Gamma + \Gamma',$$

where $\Gamma = \sum_{s=1}^{\infty} E(\Delta F_0 \Delta F_s')$ is the one-sided long-run covariance of ΔF_t . We already know that $\hat{\sigma}^2 \rightarrow_p \lambda' \Sigma \lambda$. By a similar calculation,

$$\begin{aligned}
\hat{\gamma} &= \frac{1}{T-1} \sum_{s=1}^{M-1} K(s) \sum_{t=s+1}^T (\overline{\Delta X}_t - \hat{\beta}) (\overline{\Delta X}_{t-s} - \hat{\beta}) \\
&= \bar{\lambda}' \frac{1}{T-1} \sum_{s=1}^{M-1} K(s) \sum_{t=s+1}^T (\Delta F_t - \overline{\Delta F}) (\Delta F_{t-s} - \overline{\Delta F})' \bar{\lambda} + O_p \left(\frac{M}{\sqrt{NT}} \right) + O_p \left(\frac{M}{N} \right) \\
&= \bar{\lambda}' \hat{\Gamma} \bar{\lambda} + O_p \left(\frac{M}{\sqrt{NT}} \right) + O_p \left(\frac{M}{N} \right),
\end{aligned}$$

where the remainder is negligible provided that M/\sqrt{NT} and M/N both go to zero, which is satisfied if we assume that $M/\min\{N, \sqrt{T}\} \rightarrow 0$ as $M, N, T \rightarrow \infty$. Thus, letting $\hat{\Omega} = \hat{\Sigma} + \hat{\Gamma} + \hat{\Gamma}'$ it follows that

$$\hat{\omega}^2 = \hat{\sigma}^2 + 2\hat{\gamma} = \bar{\lambda}' \hat{\Sigma} \bar{\lambda} + 2\bar{\lambda}' \hat{\Gamma} \bar{\lambda} + o_p(1) = \bar{\lambda}' \hat{\Omega} \bar{\lambda} + o_p(1) \rightarrow_p \lambda' \Omega \lambda,$$

which is identically the long-run variance of $\bar{\lambda}' \sqrt{T} \Delta \bar{F}$. Hence, $\hat{\sigma}^2$ may be replaced by $\hat{\omega}^2$.

Suppose also that in contrast to before now only the first $r_1 \geq 1$ elements of F_t are non-stationary, while the degree of integration of e_{it} is completely unrestricted. In other words, the only assumption here is that F_t contains at least one unit root.

We now show that this extension does not affect the asymptotic distribution of t_β . The reason is that the elements that are stationary are of smaller order than those that are non-stationary. Specifically, using a one to superscript subvectors corresponding to the first r_1 elements of F_t ,

$$\bar{\lambda}' \Delta \bar{F} = \begin{pmatrix} \bar{\lambda}^1 & \bar{\lambda}^2 \end{pmatrix} \begin{pmatrix} \overline{\Delta F^1} \\ \overline{\Delta F^2} \end{pmatrix} = \bar{\lambda}^1 \overline{\Delta F^1} + \bar{\lambda}^2 \overline{\Delta F^2},$$

But $\overline{\Delta F^2}$ are over-differenced, and so we get

$$\begin{aligned} \bar{\lambda}^2 \sqrt{T} \overline{\Delta F^2} &= \frac{1}{\sqrt{T}} \sum_{j=r_1+1}^r \bar{\lambda}_j \sum_{t=2}^T \Delta F_{jt} = \frac{1}{\sqrt{T}} \sum_{j=r_1+1}^r \bar{\lambda}_j \alpha_j^T F_{j0} + \frac{1}{\sqrt{T}} \sum_{j=r_1+1}^r \bar{\lambda}_j \sum_{t=0}^{T-1} \alpha_j^t \eta_{jT-t} \\ &= \frac{1}{\sqrt{T}} \sum_{j=r_1+1}^r \bar{\lambda}_j \sum_{t=0}^{T-1} \alpha_j^t \eta_{jT-t} + o_p(1), \end{aligned}$$

where $\bar{\lambda}_j$ is the j^{th} element of $\bar{\lambda}$. Moreover, since $|\alpha_j| < 1$, $\sum_{t=0}^{T-1} \alpha_j^t \eta_{jT-t} = O_p(1)$, which in turn implies

$$\begin{aligned} \bar{\lambda}' \sqrt{T} \Delta \bar{F} &= \bar{\lambda}^1 \sqrt{T} \overline{\Delta F^1} + \bar{\lambda}^2 \sqrt{T} \overline{\Delta F^2} = \bar{\lambda}^1 \sqrt{T} \overline{\Delta F^1} + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &\rightarrow_d \sqrt{\lambda^1 \Omega^{11} \lambda^1} N(0, 1) \end{aligned}$$

as $T \rightarrow \infty$, where $\lambda^1 = \lim_{N \rightarrow \infty} \bar{\lambda}^1$ and Ω^{11} is the upper left $r_1 \times r_1$ submatrix of Ω . Hence,

$$\sqrt{T}(\hat{\beta} - \bar{\beta}) \rightarrow_d \sqrt{\lambda^1 \Omega^{11} \lambda^1} N(0, 1).$$

Consider $\hat{\omega}^2$. Because of the over-differencing, Ω is singular. Specifically, by Lemma A.1 of Westerlund (2009), as $M, T \rightarrow \infty$

$$\hat{\Omega} = \begin{pmatrix} \hat{\Omega}^{11} & O_p(1/M) \\ O_p(1/M) & O_p(1/M) \end{pmatrix} \rightarrow_p \Omega = \begin{pmatrix} \Omega^{11} & 0 \\ 0 & 0 \end{pmatrix},$$

from which it follows that as $M, N, T \rightarrow \infty$ with $M/\min\{N, \sqrt{T}\} \rightarrow 0$

$$\hat{\omega}^2 = \bar{\lambda}' \hat{\Omega} \bar{\lambda} + o_p(1) \rightarrow_p \lambda' \Omega \lambda = (\lambda^1, \lambda^2)' \begin{pmatrix} \Omega^{11} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda^1 \\ \lambda^2 \end{pmatrix} = \lambda^1 \Omega^{11} \lambda^1,$$

which in turn implies that $t_\beta \rightarrow_d N(0, 1)$. Thus, t_β remains valid as long as there is at least one non-stationary factor in F_t .

On the other hand, if $r_1 = 0$,

$$\sqrt{T}(\hat{\beta} - \bar{\beta}) = \bar{\lambda}' \sqrt{T} \Delta F + \sqrt{T} \Delta e = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

which together with $\hat{\omega}^2 = O_p(1/M)$ yields $t_\beta = O_p(\sqrt{M}/\min\{\sqrt{T}, \sqrt{N}\})$. Thus, since we have assumed that $M/\min\{N, \sqrt{T}\} \rightarrow 0$, this will lead to a conservative test. Thus, for t_β not to have a degenerate limiting distribution, we need $r_1 \geq 1$. This proves the result in (9).

Appendix B: Data

Crime data

The annual crime rate data are obtained from the Swedish National Council for Crime Prevention, and are measured as the number of reported offences to the police per 100,000 of the population.²⁷ Burglary also include attempt of burglary. Theft offences constitute the largest category of crimes in terms of absolute numbers and includes shoplifting. Robbery includes both personal mugging and robbery against juristic person. Homicide includes attempt of homicide.

Macroeconomic data

The macroeconomic data include real GDP per capita, real private consumption per capita and the unemployment rate, and are obtained from the OECD database Economic Outlook, number 84.²⁸ As with the crime rates, these data are annual and cover the 1975–2008 period.

²⁷More information can be found at the web site of the National Council for Crime Prevention, <http://www.bra.se/>.

²⁸See <http://www.oecd.org/>.

Table 1: Size and power for the factor and idiosyncratic unit root tests.

N	T	α, δ	$\beta = 0$			$\beta = 1$				
			ADF ^c	ADF ^{τ}	P ^c	P ^{τ}	ADF ^c	ADF ^{τ}	P ^c	P ^{τ}
20	50	1	6.7	8.0	10.7	14.2	1.0	8.0	0.9	14.2
	100		5.9	6.4	9.1	12.3	0.6	6.4	0.2	12.3
	200		5.0	5.7	7.1	8.6	0.3	5.7	0.1	8.6
40	50		6.4	7.8	11.4	18.5	1.0	7.8	0.6	18.5
	100		6.0	6.5	9.5	12.2	0.8	6.5	0.1	12.2
	200		5.0	6.4	7.3	8.5	0.5	6.4	0.0	8.5
20	50	0.95	9.2	9.3	41.7	25.3	0.4	9.3	0.4	25.3
	100		13.4	10.3	89.8	53.1	0.0	10.3	0.0	53.1
	200		33.2	21.0	100.0	99.3	0.0	21.0	0.0	99.3
40	50		8.8	9.4	65.0	37.5	0.4	9.4	0.2	37.5
	100		14.0	10.8	99.5	75.9	0.0	10.8	0.0	75.9
	200		35.2	22.8	100.0	100.0	0.0	22.8	0.0	100.0
20	50	0.9	14.9	12.8	91.9	59.7	0.1	12.8	0.2	59.7
	100		35.2	23.4	100.0	99.6	0.0	23.4	0.1	99.6
	200		88.2	68.1	100.0	100.0	0.0	68.1	0.4	100.0
40	50		14.7	13.2	99.5	84.5	0.1	13.2	0.0	84.5
	100		35.6	23.5	100.0	100.0	0.0	23.5	0.0	100.0
	200		88.1	68.6	100.0	100.0	0.0	68.6	0.0	100.0

Notes: ADF and P refer to the Bai and Ng (2004) unit root test of the common and idiosyncratic component, respectively, where the superscripts c and τ indicate whether the test regression is fitted with an intercept or an intercept and trend. Both the lag length and the number of common factors are set equal to their true values. β refers to the trend slope, while α and δ refer to the autoregressive coefficient in the factor and idiosyncratic component, respectively.

Table 2: Size of the trend test.

N	T	$\rho = 0$			$\rho = 0.3$		
		A	B	C	A	B	C
$r_1 = 5, \delta = 1$							
20	50	7.5	7.8	5.8	11.5	10.8	15.7
	100	6.3	7.0	5.8	10.3	9.4	16.4
	200	5.4	5.7	5.1	8.2	7.4	14.8
40	50	7.3	7.9	6.0	12.5	11.5	17.3
	100	6.2	6.8	5.4	10.5	9.5	16.4
	200	5.3	5.3	4.6	8.3	7.6	15.5
$r_1 = 5, \delta = 0.5$							
20	50	7.2	7.7	5.6	11.7	10.9	15.7
	100	6.2	7.0	5.7	10.4	9.2	16.4
	200	5.6	5.8	5.1	8.1	7.5	14.6
40	50	7.2	8.0	5.9	12.4	11.5	17.3
	100	6.3	6.8	5.4	10.5	9.4	16.4
	200	5.3	5.4	4.9	8.2	7.5	15.6
$r_1 = 3, \delta = 1$							
20	50	3.3	3.7	1.1	6.3	6.3	7.1
	100	2.7	3.6	1.2	5.5	5.9	7.0
	200	2.7	2.9	0.9	4.8	4.6	5.9
40	50	3.6	3.9	1.7	6.9	6.6	7.3
	100	2.9	3.8	1.3	5.7	5.7	7.0
	200	2.9	3.0	0.9	5.2	4.9	6.4

Notes: ρ refers to the first-order autoregressive serial correlation coefficient of the factors, δ refers to the autoregressive coefficient of the idiosyncratic component, and r_1 refers to the number of unit roots among the five factors. The autoregressive coefficient in the stationary factors is set to 0.5. Columns A, B and C indicate whether the bandwidth has been set as a function of T , by using the Newey and West (1994) rule or set equal to zero.

Table 3: Size and power for the sequential unit root tests.

N	T	δ	ADF ^c	ADF ^{τ}	ADF ^{seq}	P^c	P^τ	P^{seq}
$\alpha = 1$								
20	50	1	6.7	11.9	8.0	10.7	14.2	11.1
	100		5.9	10.1	6.4	9.1	12.3	9.5
	200		5.0	8.9	5.7	7.1	8.6	7.5
40	50		6.4	11.6	7.8	11.4	18.5	12.2
	100		6.0	10.5	6.5	9.5	12.2	9.8
	200		5.0	9.4	6.4	7.3	8.5	7.3
20	50	0.95	6.8	11.9	8.0	41.9	25.8	39.9
	100		6.1	10.2	6.4	89.1	53.1	85.2
	200		5.5	9.4	5.8	100.0	99.3	99.9
40	50		6.3	11.6	7.7	64.7	37.7	61.3
	100		5.9	10.3	6.4	99.3	76.1	96.5
	200		5.0	9.5	6.3	100.0	100.0	100.0
$\alpha = 0$								
20	50	1	99.8	100.0	100.0	9.0	11.8	11.8
	100		99.9	100.0	100.0	8.6	10.9	10.9
	200		100.0	100.0	100.0	6.9	8.4	8.4
40	50		100.0	100.0	100.0	9.8	15.0	15.0
	100		100.0	100.0	100.0	8.6	10.5	10.5
	200		100.0	100.0	100.0	7.1	7.9	7.9
20	50	0.95	100.0	100.0	100.0	39.6	21.8	21.8
	100		100.0	100.0	100.0	89.5	51.2	51.2
	200		100.0	100.0	100.0	100.0	99.3	99.3
40	50		100.0	100.0	100.0	61.8	32.5	32.5
	100		100.0	100.0	100.0	99.5	73.7	73.7
	200		100.0	100.0	100.0	100.0	100.0	100.0

Notes: ADF^{seq} and P^{seq} refer to the Bai and Ng (2004) unit root test of the common and idiosyncratic component, respectively, based on the sequential procedure. See Table 1 for an explanation of the remaining features.

Table 4: Cross-county correlations.

Test	Burglary	Theft	Robbery	Homicide
Average correlation	0.98	0.99	0.97	0.85
CD	82.57	83.82	81.64	72.23
<i>p</i> -value	0.00	0.00	0.00	0.00

Notes: The results are for the demeaned first differenced series. The CD statistic tests the null of no cross-correlation. The *p*-values are from the asymptotic normal distribution.

Table 5: Unit root tests of the estimated factors.

Crime	Factor	Constant			Trend		
		var	$\hat{\alpha}$	ADF ^c	var	$\hat{\alpha}$	ADF ^{τ}
Burglary	1	30.89	0.96	-0.41	31.13	0.78	-1.81
	1	29.30	0.92	-1.73	29.77	0.81	-1.65
Theft	2	22.24	0.91	-1.37	22.10	0.49	-3.23
	1	34.18	0.94	-1.27	34.61	0.65	-2.56
Robbery	2	14.93	0.50	-3.33	15.08	0.50	-3.30
	3	10.60	0.70	-1.96	10.73	0.40	-3.76
Homicide	4	9.86	1.08	1.04	9.94	0.85	-0.98
	5	7.46	1.01	0.14	7.52	0.66	-2.14
	1	20.66	0.84	-0.79	20.70	0.53	-1.79

Notes: $\hat{\alpha}$ refers to the estimated first-order autoregressive coefficient, while var refers to the percentage share of the total variance in the data. The ADF test is superscripted by c or τ to indicate whether a trend has been included. The 5% critical values for the model with and without a trend are given by -2.864 and -3.409 , respectively. The lag length is determined using the Schwarz Bayesian criterion. The number of factors is determined using the IC_2 criterion of Bai and Ng (2002).

Table 6: Tests for the presence of a trend.

Crime	$\hat{\beta}$	$ t_{\beta} $	p -value
Burglary	-11.51	0.92	0.36
Theft	32.72	2.46	0.01
Robbery	1.27	2.34	0.02
Homicide	0.16	3.21	0.00

Notes: $\hat{\beta}$ refers to the estimated trend slope with $|t_{\beta}|$ being the associated double-sided t -statistic for the null of a zero slope. The p -value is based on the normal distribution.

Table 7: Panel unit root tests of the estimated idiosyncratic component.

Crime	Constant		Trend	
	P^c	p -value	P^T	p -value
Burglary	6.00	0.00	11.66	0.00
Theft	4.01	0.00	8.52	0.00
Robbery	3.73	0.00	9.76	0.00
Homicide	9.86	0.00	24.84	0.00

Notes: P^c and P^T refer to the Bai and Ng (2004) test with a constant and a constant and trend, respectively. The p -values are based on the normal distribution, and the lag length is determined using the Schwarz Bayesian criterion.

Table 8: Estimated factor regressions.

Explanatory variable	Burglary		Theft		Robbery					Homicide
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 1
Lag dependent variable		0.00					0.00	0.00	0.00	0.00
Unemployment	0.35	0.11	0.16	0.11	0.14	0.00	0.06	0.70	0.06	0.03
Private consumption	0.71	0.04	0.27	0.04	0.24	0.20	0.35	0.21	0.22	0.04
GDP	0.15	0.59	0.12	0.59	0.39	0.00	0.11	0.28	0.26	0.42
R^2	0.10	0.41	0.71	0.41	0.12	0.41	0.11	0.08	0.36	0.73
F	0.40	0.01	0.00	0.01	0.28	0.00	0.32	0.46	0.01	0.00

Notes: The table reports the p -values for the exclusion restriction of both the contemporaneous and lagged values of the relevant explanatory variable. An unreported constant and the contemporaneous value of each regressor are always included, while the number of lags is determined using the Schwarz Bayesian criterion. The F -statistic is for the null hypothesis that all the explanatory variables but the constant can be excluded.